Time-of-day effects in icon usability.

Tyrer, Victoria

TIME-OF-DAY EFFECTS IN ICON USABILITY

Victoria Tyrer

Thesis submitted for the degree of Doctor of Philosophy

University of Wales, Swansea

July 2003

Table of Contents

Acknowledgements

Firstly I must thank my supervisors Professor. Simon Folkard and Dr. Sine McDougall for all their invaluable help and advice throughout the last three years and indeed throughout my undergraduate years. Also many thanks to Neil Carter, Philip Heard and Peter Matthews for their technical support. Thanks to all my research participants for sitting through hours of sometimes gruelling experimental sessions and I cannot forget Sam, Sarah, Meryl, Alison and Jo who have always been more than happy to share the moans and groans! Above all, I must, of course, thank Mum & Dad for their unconditional love and support throughout the highs and lows of my education and for blessing me with the genes (and money!) that have helped me to achieve so much. Without these people this thesis would never have been possible.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.. ....(candidate)

Date..25th July 2003..........................

Statement 1

This thesis is the result of my own investigations, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed. ....(candidate)

Date..25th July 2003.......................

Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed. ....(candidate)

Date..25th July 2003......................

Abstract

To date there appears to have been only one study that has examined time-of-day effects in icon interpretation. McFadden and Tepas (1997) found the time taken to respond to iconic stimuli to vary according to the time-of-day and found the exact time-of-day trend to vary according to the memory load involved in the task. This study was replicated, using slightly modified stimuli, and similar findings were obtained to McFadden and Tepas' earlier study.

A series of experiments subsequently examined the effects that different icon characteristics and other changes in task demands had on the observed time-of-day trends. The first of this series, compared icons that were made up of a series of features (multi-feature) with those that were relatively wholistic (gestalts). Gestalt icons were found to markedly improve usability by dramatically reducing response times. Additionally, a trend was noted for the exact timing of peak performance to vary slightly according to icon type, with the multi-feature icons showing a slightly earlier peak in performance.

The experiments that followed used icons that had been varied orthogonally in terms of their complexity and concreteness and examined other variations in task demands in terms of the semantic memory component required, the visual memory component involved, the difficulty of response required and the difficulty of icon discrimination. Results suggested that icon tasks requiring semantic memory were not susceptible to time-of-day effects. Similarly, neither differences in icon discrimination nor visual memory were critical in determining the diurnal trend observed. Surprisingly, it was difficulty of response that appeared to be a critical factor in consideration of the influence of exact task demands in icon search tasks. Interestingly however, it appeared that abstract, rather than concrete, icons may show more pronounced diurnal performance trends.

It was proposed that the effects of different task demands on the observed time-of-day trends exerted their effects through their influence on working memory load, with higher memory load tasks showing an earlier performance peak relative to lower memory load tasks. A framework was proposed for the understanding, and development of, these time-of-day effects in icon usability.

Chapter 1
Circadian Rhythms

1.1. Introduction

Rhythms are an inherent part of life for all living organisms and can be detected in organisms as diverse as unicellular creatures and man (Wever, 1979). All organisms show rhythmicity in various biological functions – "rhythmicity is a ubiquitous biological phenomenon" (Wever, 1979 p. 1). Periodicity in nature expresses itself in the annual seasons, phases of the moon, day and night earth rotation and the changes in the tide. Some or all of these influence the activity of biological fluctuations (Kleitman, 1939). "There is apparently no organ and no function in the body which does not exhibit a similar daily rhythmicity" (Aschoff, 1965 p. 1427). Here, attention will primarily focus on the periodicity found in man.

Kleitman (1939) distinguished between rhythms and cycles, defining a rhythm as "a regularly recurring quantitative change in some particular variable biological process...two conditions are necessary to make such a recurring change into a rhythm: (a) it must be extrinsic in origin, depending upon a regular change in the environment; (b) when fully established it must persist for some time, even when the environmental changes are absent." (p.131). A cycle is defined as: "a repetitive series of events or...successive changes of state, either qualitative or quantitative in nature...its distinctive feature is the order of occurrence, rather than duration. Cycles are intrinsic in origin and have to run their course to be completed. They may be influenced by internal and external conditions, which may affect them quantitatively, but seldom qualitatively e.g. the cardiac cycle." (p.131). Kleitman (1939) concluded "the development and maintenance of the 24 hour sleep-wakefulness and body temperature rhythm stem from being born into and living in, a family and community run according to alternations of light and darkness, resulting from the period of rotation of the earth around its axis." (p.147).

Probably the most commonly known rhythm is the circadian rhythm (from the Latin: 'circa' = about; 'dian' = day), relating to rhythms with approximately a 24-hour cycle. The majority, if not all, biological functions have a circadian rhythm – "there is hardly a tissue or function that has not been shown to have some 24hr variation" (Aschoff and Wever, 1981 p.311). It seems that the existence of this biological phenomenon is of great importance to our well-being - "the pervasive nature of such rhythmic components in physiology and behaviour suggests that this circadian...temporal organization is vital to the overall well-being of the organism" (Campbell, 1992). Core body temperature, hormonal secretion, rest and activity, sleep and wakefulness are all biological functions that have been found to fluctuate over the day (Campbell, 1992).

Due to the predominance of rhythms that have a duration of about one day, other physiological rhythms with shorter or longer frequencies are termed as ultradian or infradian respectively (Campbell, 1992). Rhythms which demonstrate a cycle of less than one day, for example the sleep stages are arranged around a 90-minute cycle of REM and slow-wave sleep (Carlson, 1995), are termed ultradian rhythms. Those rhythms that have a cycle of quite long duration are known as infradian rhythms and include the menstrual cycle, for example.

Here, we are concerned with circadian rhythms in human psychological processes. This is of particular importance because if there are rhythmical processes within a subject who is performing a task, then "a stimulus at one time will not have the same effect as the same stimulus at another time" (Oatley and Goodwin, 1971 p.2). Oatley et al (1971) stated that some rhythms could be fundamental to performance. They believe that good performance and good experimental control can only be attained by co-operating with rhythms.

Indeed, literature to date has suggested that variations across the day in the efficiency and accuracy with which we perform particular tasks do exist. In fact, as Campbell (1992) highlighted, most measures of performance show approximately 24-hour oscillations, although shorter-term fluctuations have been suggested as well. For instance, the human sleep-wake cycle profoundly affects performance and is mainly a circadian rhythm but contains ultradian components.

The aim of this thesis was to examine the influence of the time-of-day, and therefore the stage of the circadian rhythm, on human performance in icon search tasks. This research is important to those working within the transport and information technology industries, since it addresses such questions as: Can we predict the effectiveness of icons according to the time-of-day they are viewed? Does the time-of-day at which individuals work with the aid of a visual display (for example, air traffic controllers, pilots, drivers) influence the likelihood of an accident due to human error?

The remainder of this chapter will be concerned with general information about human circadian rhythms and then with the influence these rhythms have in general areas of human performance. Later specific consideration will be given to the influence of circadian rhythms on human memory, before examining the arousal explanation and then more contemporary theories of diurnal performance rhythms. Finally, the limited research that has been conducted to date on the influence of time-of-day on the completion of icon search tasks will be considered.

1.2. Human Circadian Rhythms

On earth, life has evolved in an environment subject to changes produced by planetary movements. The earth rotates on its' axis and this results in the 24 hour light/dark cycle. The earth's rotation around the sun causes seasonal changes in light and temperature (Folkard, 1996). The periodicity of circadian rhythms corresponds to these rhythms in the environment, thus it appears that these rhythms contain internal and external components (Wever, 1979).

Oatley et al (1971) suggested that many biological and behavioural rhythms have originated due to adaptation to these environmental cues. They point out that there has been much debate over whether rhythmicity is generated by the organism and becomes entrained to the environmental cues or whether the external rhythm produces the rhythm in the organism. It has, however, been established that at least some circadian rhythms are produced from within the organism regardless of any rhythms in the environment.

Oatley et al (1971) continued, that organisms often need to predict the external environment, thus it is something that the brain has to model. This model would consist of endogenous oscillators, the ideal example of which are the circadian rhythms that have oscillators modelling the sequence of day and night. Endogenous clocks serve to wake the organism ready for a period of activity. The endogenous clock is responsible for the timing, while the light and darkness in the environment keeps the clock synchronised with the rhythms produced by the earth's rotation. Folkard (1996) supported this notion proposing that during evolution environmental changes became internalised to allow organisms to anticipate the changes. He believed this ability has an adaptive value for most organisms and has probably been strengthened through natural selection.

The notion that rhythms in the environment have become internalised and are paralleled by some sort of endogenous clock, which is then kept in synch by environmental cues, is by far the most appealing explanation for how circadian rhythms came to exist. This notion also receives support from temporal isolation studies, which will be discussed later in this chapter. In the meantime consideration will be given to research trying to establish the location and nature of this 'internal clock'.

1.2.1 The Nature of the 'Internal Clock'

In the rat the main biological clock is found in the Suprachiasmatic Nucleus (SCN) of the hypothalamus. Lesions here disrupt many cycles (Carlson, 1995) and have been found to cause circadian arrhythmia in both hamsters and rats (Moore and Eichler, 1972; Stephan and Zucher, 1972). However, Inouye and Kawamura (1979) recorded circadian rhythmicity in many other brain sites in addition to the suprachiasmatic nuclei. Yet when the suprachiasmatic nuclei were surgically isolated rhythmicity in the rest of the brain was lost, while rhythmicity remained in the isolated suprachiasmatic nucleus. This suggests that the suprachiasmatic nucleus was autonomous from the other 'clocks' that were controlled by it. Indeed, Pittendrigh (1981) discussed circadian rhythmicity in terms of the suprachiasmatic nucleus being the overall pacemaker, controlling temporal events in peripheral organs such as the liver. Further evidence suggests, that the "ticking" of the clock in the suprachaismatic nucleus, equivalent to the pendulum in a mechanical clock, is probably intrinsic to each neuron, such that each has it's own clock (Carlson, 1995).

But how does the suprachiasmatic nucleus exert its' controlling influence? It has been found that the group of neurons surrounding the capillaries that serve the suprachiasmatic nucleus contain rough endoplasmic reticulum, typical of neurosecretory cells. Therefore the suprachiasmatic nucleus may gain some of its' control through the secretion of neuromodulators (Carlson, 1995).

It would seem that the pacemaker *alone* is entrained by zeitgebers, which in turn then entrain and control the "slave oscillations" (Pittendrigh, 1981 p. 70). In support of the notion that light is the main zeitgeber ('time-giver') for most organisms, fibres projecting from the retina have been found in the suprachiasmatic nucleus (Carlson, 1995). It has been suggested that light hitting the retina results in messages being sent to the suprachiasmatic nucleus which then sends messages to suppress the secretion of melatonin by the pineal gland (Folkard, 1996). Melatonin production suppresses alertness and therefore aids sleep, thus conversely the suppression of melatonin secretion induces wakefulness.

Although this only briefly reviews research into the nature of the internal clock(s), it is clear that the control of circadian rhythms in man is a complex chain of events involving the interaction of neurons, neuromodulators, hormones and other organs of the body. However, overall control seems to be exerted by the suprachiasmatic nucleus.

1.3. Specific Biological Rhythms

1.3.1 The Temperature Rhythm

Here interest is in the measurement of temperature, a physiological variable, as a reflection of the activity of endogenous circadian rhythms, against which any changes in performance may be compared (Hockey and Colquhoun, 1972). Despite considerable controversy as to its' usefulness, variation in body temperature has been measured and used for many years in the study of time-of-day effects in performance measures. Indeed the 24-hour variation in body temperature has been discussed as far back as 1842 (Kleitman, 1939). Temperature has been used as a reflection of the stage of an individual's circadian rhythm for a number of reasons. It is one of the most observable circadian rhythms (Colquhoun, 1971), is easy to measure, is one of many biological functions that exhibits a circadian rhythm having a rather constant periodicity of 24 hours (Hockey and Colquhoun, 1972) and is not influenced by external variables such as changes in activity level. Even with bed rest the fluctuation in deep body temperature remains rhythmic (Wever, 1979; Kleitman, 1939).

As these latter points suggest, the temperature rhythm is regarded as autonomous or free-running. Certainly, left to run-free the temperature rhythm runs on a 25 hour cycle (Folkard, 1996; Campbell, 1992), and thus has a strong endogenous component (Wever, 1979). The mean 25 hour rhythm only deviates between individuals by about half an hour (Wever, 1979). Typically, the temperature rhythm and other rhythms (for example, the sleep-wake cycle) remain synchronized to each other (Campbell, 1992).

Temperature reaches a maximum at about 2000-2100 hours and falls to a minimum (nadir) at about 0400-0500 hours (Hockey and Colquhoun, 1972; Smith, 1992). Accordingly, temperature ranges from about 36.2 °C early in the morning to about 36.9°C in the evening at about 2000 hours (Folkard, 1983). Major sleep episodes normally occur at approximately the nadir of the temperature rhythm and end many hours after the nadir has occurred (Campbell, 1992).

As temperature, and many other, physiological rhythms fluctuate over the day, it seems inevitable that mental efficiency will also fluctuate as it seems fair to assume that the brain is supported by at least some of these physiological functions that show variations over the day (Colquhoun, 1971). This seems reasonable especially when consideration is given to the fact that part of the brain, the hypothalamus, is involved in temperature control.

It has long been established that temperature is controlled by the hypothalamic brain structure. For instance, Kleitman (1939) found the hypothalamus in cats to be warmer than the cortex. Activity and excitement increased cerebral temperature (cortical and hypothalamic) while rest produced a drop in temperature. Sleep produced a drop in temperature that was greater in the hypothalamus than in the cortex. When temperature did not drop in the hypothalamus, it failed to drop anywhere. On awakening the rise in temperature occurred first in the hypothalamus. Often during sleep the temperature of the hypothalamus would begin to rise, then the animal would change position and settle and the hypothalamic temperature would begin to decrease. As a result, it was concluded that the hypothalamus was less active during sleep than wakefulness, due to its temperature changes. Kleitman (1939) believed that the greater fall in temperature in the hypothalamus was not due to it being further away from the surface, as other regions that are at an equal depth did not mirror the drop. Neither were changes in circulation believed to be responsible for the changes in temperature.

Kleitman (1939, 1963) actually found a parallelism between the body temperature rhythm and the diurnal trend seen in reaction times. As a result Kleitman (1939, 1963) believed there was a causal relationship between performance and temperature. Subsequently however, this notion has been rejected (Colquhoun, 1971). Indeed more recently, Smith (1992) highlighted contemporary thinking in this area stating that although performance efficiency may parallel the rise and fall in the body temperature curve on occasions, there are certainly exceptions to this rule and it is important to realise that this relationship is by no means a causal one. There is however, overwhelming evidence that performance levels fall to a minimum at night when body temperature is at its' lowest (Smith, 1992). Evidence that the link between the temperature rhythm and other performance rhythms is by no means clear cut is provided by Owens, MacDonald, Tucker, Sytnik, Totterdell, Minors, Waterhouse and Folkard (2000). They highlighted the fact that studies using serial visual search speed have claimed a parallel between performance and core body temperature, a finding supported by Folkard (1996). However, Owens et al (2000) subsequently concluded that overall, performance variables that showed diurnal variations had trends that failed to parallel the trends seen in core body temperature. Similarly, Folkard (1996) pointed out that both performance on memory-loaded tasks and alertness ratings have rhythms that behave differently to those in body temperature. Further, as alertness and temperature rhythms behave differently (Folkard and Monk, 1985), it has been concluded that temperature cannot be used as an indication of alertness (Owens et al, 2000). Yet, Monk et al (1997) found subjective alertness to correlate with performance equally as well as body temperature. This complex relationship between temperature and psychological and performance measures is considered in more detail later in this chapter.

Carrier and Monk (2000) considered the possibility that other physiological variables could be used as indices of the stage of the circadian rhythm and therefore performance efficiency. They stated "endogenous circadian performance rhythms are controlled by the same pacemaker that drives the endogenous circadian rhythm of body temperature" (p.174), further they stated "this pacemaker also drives a number of physiological rhythms, including plasma cortisol and plasma melatonin" (p.174-726). Thus any one of these may be used as an index of the stage of the circadian rhythm rather than temperature (Carrier and Monk, 2000). Certainly, Monk, Buysee and Reynolds (1997) found that temperature and cortisol rhythms correlated with performance measures, although melatonin did not correlate so well. Positive correlations were found between good performance and high temperature values and good subjective alertness, while negative correlations were found between good performance and high plasma cortisol levels and plasma melatonin levels. However, the parallelism between these physiological variables and performance was not strong therefore Carrier and Monk (2000) concluded that caution should be exercised in extrapolating from levels of one to levels of the other. Indeed, Carrier and Monk (2000) believed that it is best to consider performance rhythms as being independently controlled by the circadian timing system and time spent awake (see section 1.9), resulting in a trend that sometimes happens to coincide with those trends seen in some physiological variables "without necessarily being directly mediated by any particular physiological rhythm" (p.726).

1.3.2 The 'Sleepiness' Rhythm

Not all rhythms are endogenously controlled, while temperature runs on a 25 hour cycle when left to run-free, the sleep-wake cycle may change to as much as 36 hours (Folkard, 1996). This is taken as evidence that the sleep-wake cycle is relatively exogenous in origin (Folkard, 1983). Although, despite this apparent flexibility of the sleep-wake cycle, Kleitman (1963) found evidence to conclude that it was unaffected by changes for example, in the seasons.

Humans have the tendency to obtain their daily sleep in one long episode, consequently the human sleep system has been considered as being monophasic. Now however it has been postulated that this is not the case and that human sleep should be viewed as polyphasic with a minimum of two preferred phases for sleep within 24 hours (Campbell, 1992). When subjects are kept in a time-free environment and are instructed to eat and sleep whenever they like, major sleep episodes occur, that is sleep propensity is highest, at about the same time as the nadir in the core body temperature rhythm (Lack and Lushington, 1996; Campbell, 1992), but also there is the tendency for a second period of sleep to occur when core body temperature is higher. The sleep periods occurring around the nadir last approximately 8 hours, while those occurring when core body temperature is higher last about 1 ½ - 2 hours. Hence under entrained conditions longer sleep periods correspond to long nocturnal sleep sessions and the shorter sleep periods correspond to afternoon 'nap' sessions (Campbell, 1992). Paradoxically, sleep propensity is at its' lowest about 8 hours before the minimum temperature, that is, when most people usually retire to bed (Lack and Lushington, 1996).

Immediately following awakening from a night's sleep, a 'sleep inertia' effect has been observed where performance decrements often accompanied by dysphoria and confusion occur. Interestingly, cognitive performance has been found to recover from sleep inertia more quickly than motor performance (Ferrara, De Gennaro and Bertini, 2000). Generally the effects of 'sleep inertia' are not long lasting, usually persisting for less than 20 minutes in individuals who have rested well. However, the effects of sleep inertia can be more severe in those who have suffered sleep deprivation (Campbell, 1992) and in those who have been awakened from stage 4 sleep (Dinges, Orne, Evans and Orne; 1981). Similarly, the more slow wave sleep achieved in the preceding sleep period, the more severe the sleep inertia effect will probably be, thus the composition of the prior sleep period is of importance (Dinges, Orne and Orne, 1985; Naitoh and Angus, 1989). Furthermore, Campbell (1992) stated that circadian influences are also likely to affect the amount of sleep inertia suffered.

1.3.3 The Post-Lunch Dip

Our propensity to fall asleep is known to vary throughout the day, this occurs irrespective of our level of activity. Many researchers have found evidence of a dip in performance during the early afternoon. Colquhoun (1971) stated that the feeling most commonly reported during this post-lunch period, by those who exhibit a drop in performance, is that of sleepiness. Accordingly, Lenne, Triggs and Redman (1997) found performance dips at 1400 hours for all of their performance measures and highlighted the fact that these performance dips corresponded with the increase of car accidents in the early afternoon. They did, however, report that despite this increase in sleep propensity in the early afternoon, subjective alertness often peaks at 1400 hours. Note this is contrary to Colquhoun's (1971) statement that people report sleepiness at this time.

So what may be the cause of this post-lunch dip? Colquhoun (1971) proposed that we equate 'sleepiness' with 'arousal' and argued that the general level of sleepiness falls, that is arousal rises, during the course of the waking day, reaching a minimum in the evening. Subsequently the sleepiness level rises again until sleep occurs. Colquhoun (1971) also believed that we should suppose a temporary increase in sleepiness level, that is a decrease in arousal, occurs at about the time at which we take lunch. Colquhoun (1971) suggested that this temporary rise in sleepiness might exist due to a secondary cycle that once served a biological purpose but no longer does, being superimposed upon the circadian cycle. Kleitman (1963) proposed that such fluctuations in arousal reflect the underlying REM 90 minute cycle persisting into the waking period. More recently, Broughton (1998) suggested that the post-lunch dip is due to an increase in homeostatic sleep propensity being over-powered by a circadian arousal component that will peak later in the evening.

Interestingly, Colquhoun (1971) noted that the temperature rhythm does not exhibit this post-lunch dip and believed that body temperature should not be considered as an index of the level of arousal over the day except on occasions where changes in temperature happen to coincide with fluctuations in arousal. However, this proposal is clearly circular. More convincing is Hockey and Colquhouns' (1972) belief that the absence of a post-lunch dip in the temperature rhythm is evidence that temperature is not the only determinant of performance.

An important point to note here is that just as the temperature rhythm is not related to physical activity, so the post-lunch dip is not related to the ingestion of food (Wever, 1979). Indeed, Folkard and Monk (1985) believed that the best evidence for the existence of an endogenous component in the post-lunch dip comes from isolation studies where a dip in performance at about this time can be seen even though the performance rhythm and the food ingestion rhythm are running on different periods. In support of this, Hockey and Colquhoun (1972) described an unpublished study by Blake where the timing of lunch was varied, no difference in performance was observed when the timing of lunch was changed from 1200 to 1400. Hence we seem to be dealing with a genuine dip in performance during the post-lunch period that is unaffected by external factors. Conversely however, studies have found that the post-lunch dip can be made worse by a lunch that is high in carbohydrate (Craig, Baer and Diekmann (1981). Similarly, Smith and Miles (1986b, 1987b) found that certain impairments in performance, in particular, impairments in sustained attention, depend on the ingestion of a lunchtime meal. Consequently, it seems that the appearance of a post lunch dip associated with the ingestion of food is dependent upon the nature of the task. Also there is the suggestion that the influence of lunch on performance depends on the nature of the meal (Smith, 1988). Furthermore, there is evidence that the post-lunch dip can be removed with caffeine (Smith, Rusted, Eaton-Williams, Savory and Leathwood, 1990) and noise (Smith and Miles, 1986a). Therefore, there is evidence that the post-lunch dip is flexible, in that it can sometimes be influenced by various external factors.

Contrary to much research reporting the occurrence of this post-lunch dip, some studies have failed to find evidence of a fall in performance during the post-lunch period (for example, Hughes and Folkard, 1976; Christie and McBearby, 1979; Folkard and Monk, 1987). Perhaps individual differences hold the answer. Lavie and Segal (1989) found a much clearer post-lunch dip for morning types than evening types when using an ultrashort sleep/wake paradigm, whilst Monk et al (1996) found participants who showed a post-lunch dip in performance had a 12 hour component of rectal temperature that showed a higher amplitude and a later peak than those participants who showed no evidence of a post-lunch dip in performance. Further, it has been found that people with low levels of anxiety exhibit a greater post-lunch dip (Craig, Baer and Diekmann, 1981, Smith and Miles, 1986a). Thus there is evidence that the post-lunch dip can be exacerbated or improved by individual or external influences. Clearly, establishing a way to overcome the post-lunch dip, when it occurs, would hold many advantages. Consequently, brief consideration will now be given to the influence of caffeine on periods of sub-optimal performance.

1.3.4 Overcoming the Post-Lunch Dip & Optimising Performance

Caffeine crosses the blood-brain barrier with ease and is rapidly absorbed. If caffeine can improve performance and prevent accidents then the advantages of its' use are obvious. If, on the other hand, it lacks any beneficial effects then there is no need for its' ingestion (Lieberman, 1992). Justification for why a compound such as caffeine may help overcome decrements in performance resulting from circadian deficits comes from early research. Kleitman (1939) found changes in the body temperature curve after consuming alcohol or caffeine before going to bed. Caffeine doses of 260-390mg caused a rise in temperature, while smaller doses of 130mg exerted no influence.

Perhaps then it is not surprising that it has been well established that caffeine increases alertness, consequently caffeine has been used in low alertness situations. Caffeine has been shown to have advantageous effects when a person is deprived of sleep (Bonnet and Arand, 1994) and, when coupled with a nap, to drivers needing to overcome sleepiness (Horne and Reyner, 1996). In support of Horne et al (1996), it has also been found that the combination of nap and caffeine ingestion is most effective in maintaining nocturnal alertness (Bonnet and Arand, 1994). Indeed, caffeine has been shown to be of benefit after lunch when the post-lunch dip takes its toll on performance (Smith et al, 1990; Smith, Rusted, Savory, Eaton-Williams, and Hall, 1991), and has been found to counteract test session fatigue (Smith, Clark and Gallagher, 1999) while caffeinated-beverage *deprivation* has been found to result in increased fatigue (Lane, 1997). Further, Hindmarch, Quinlan, Moore and Parkin (1998) found that over the complete day, consuming tea instead of water and caffeinated rather than decaffeinated beverages, prevented the decline in alertness and cognitive capacity seen with water consumption. In support of this, Lieberman (1992) stated that even in low to moderate doses caffeine seems to enhance alertness and reduce fatigue. Yet conversely, Linde (1995) investigated the psychological effects of caffeine in both fatigued and non-fatigued individuals and found that overall the influence of caffeine on performance and subjective fatigue were weak.

Caffeine has also been found to be beneficial even when fatigue is not a factor. For instance, beneficial effects of caffeine have been observed in reaction time and vigilance tasks (Lieberman, 1992). To expand, caffeine has been found to exert a positive influence on simple and choice reaction times and has advantageous effects in responsiveness to auditory and visual stimuli, such effects are likely to be related to the maintenance of vigilance (Lieberman, 1992). Interestingly, Lieberman (1992) highlighted that high doses of caffeine may be less advantageous on reaction times than moderate doses. Furthermore, Miller, Lombardo and Fowler (1995) found a discrimination task to be influenced by a combination of time-of-day, dosage of caffeine and their interaction. Also, caffeine has been found to be beneficial in perceptual-motor tasks and sustained attention tasks both at night and also during the day (Smith, Brockman, Flynn and Maben, 1993) and caffeine has been found to increase the speed of encoding of new information (Smith et al, 1999; Smith, 2000; White, 1998).

However, although caffeine appears to moderate time-of-day effects, several extraneous variables need to be accounted for when considering its effects. These include dosage, time of administration, function under examination, impulsivity of the individual (Smith et al, 1991), user history (Mitchell and Redman, 1992; Lieberman, 1992), morningness (Ryan, Hatfield and Hofstetter, 2002) and personality (Lieberman, 1992).

1.4. The Desynchronisation of Rhythms

Considering the motor activity of the rat, Carlson (1995) found that if the dark-light cycle is shifted by six hours then the animal's activity periods will follow, indicating that they must be controlled by external (exogenous) cues. Yet, if dim lights are left on continuously, the cycle in the rats' activity remains, indicating that the rhythm is controlled endogenously. Then again, when the rat's clock was left to "run-free" the animal began its activity periods about one hour later each day. This phenomenon is typical of most organisms. Accordingly, most organisms demonstrate a cycle of a little longer than 24 hours when their internal clocks are left to run free. Daily changes in illumination serve to keep the clock adjusted to 24 hours. When human internal clocks are left to run free, most people live "days" that are about 25 hours long, supporting the notion that circadian rhythms are controlled endogenously but are normally 'guided' by the environmental rhythms.

Folkard (1996) turned the focus from animals to humans. He believed that the best evidence that human circadian rhythms are controlled by an endogenous body clock comes from studies where people have been isolated from the environment and zeitgebers. In these temporal isolation studies, a subject is placed in an environment away from external time cues, thus allowing their rhythms to 'run-free' (Aschoff and Wever, 1981). Aschoff and Wever (1962) who isolated participants from all zeitgebers for up to nineteen days, and Siffre (1964) who lived underground in a cave and therefore was also isolated from zeitgebers for two months, found that all physiological processes ran to an approximate 25-hour rhythm. Aschoff and Wever (1981) and Wever (1979) stated that results of temporal isolation studies have shown that while the rhythms run to a slightly longer period of about 25 hours they still usually remain synchronised with each other, as they do under normal conditions where the circadian rhythms of man run in a phase-relationship to the sleep-wake schedule which is maintained by zeitgebers. Thus, the existence of day-to-day flexibility of the circadian system within individuals is well established (Webb and Agnew, 1978), at the level of the individuals' behaviour, the human circadian system appears somewhat 'sloppy' (Campbell, 1984; Campbell and Zulley, 1988). It is this flexibility of the human circadian system that allows the free-running endogenous rhythms to become entrained (or synchronised) by environmental, social and behavioural cues. However, erratic exposure to environmental time cues can cause problems with the synchronisation of rhythms to the surrounding world, for instance it has been shown that the circadian rhythm of temperature does not always synchronise to the erratic environmental information that flight crewmembers can be exposed to, thus the flight crewmembers showed evidence of desynchronisation (Gander, Gregory, Graeber, Connell, Miller, Rosekind and Mark, 1998).

1.4.1 Internal Desynchronisation

Wever (1979) believed the fact that human circadian rhythms run autonomously when external cues are removed demonstrates that circadian rhythms are of endogenous origin. Under such conditions Aschoff and Wever (1981) and Wever (1979) illustrated that it is possible for some individuals to exhibit a state of internal desynchronisation – "a state where different variables oscillate with different periods" (Wever, 1979 p.43). More specifically, this is where the temperature, and some other, rhythms run at different periods to the sleep-wake cycle.

For most lower animals, it is the light dark cycle that is the most powerful zeitgeber. For humans the story is a little different. While the sleep-wake cycle follows changes in a zeitgeber such as light, circadian rhythms in body temperature and in other physiological functions have a more restricted range of entrainment and under normal levels of artificial illumination they can only be entrained down to a period of 23 hours or up to 27 hours. Outside this range, rhythms that are mainly controlled by the endogenous body clock 'break out' from the sleep-wake cycle and free run with their endogenous period (Folkard, 1996). Aschoff and Wever (1981) stated that it is normal for the temperature rhythm to break away first. Therefore by using artificial 'day' lengths of less than 23 hours or more than 27 hours, internal desynchronisation can be induced in everyone (Folkard, 1996). Wever (1979; 1982) reported that when spontaneous internal desynchronisation occurred in his subjects the temperature rhythm kept to a period of approximately 25 hours while the sleep-wake schedule kept to a longer period of 30-40 hours or a shorter period of 15-20 hours.

An important distinction to note here is between real internal desynchronisation and apparent internal desynchronisation (Wever, 1979). Internal synchronisation occurs where different rhythms run temporally constant to each other. In some subjects their 25 hour temperature rhythm is combined with an activity period that is double or half the temperature value but the two rhythms are internally synchronised with each other, so this looks like internal desynchronisation due to the different periods of the rhythms. This state is known as apparent internal desynchronisation (Wever, 1979).

1.4.2 External Desynchronisation

Folkard (1983) described another related phenomena known as external desynchronisation. This is where a conflict occurs between an individual's 24-hour rhythms and external zeitgebers. This occurs when a person is forced to adapt to an unusual sleep-wake schedule, much like shiftworkers are forced to do when working a rotating schedule involving day and night work.

1.4.3 Fractional Desynchronisation

Another related phenomenon is 'fractional desynchronisation'. When a subject whose rhythms are running-free is exposed to a strong zeitgeber, for example absolute periods of darkness signalled with a gong, and the subject is thus forced to adopt a specific sleep-wake schedule the periodicity of which is slowly shortened or lengthened, the temperature rhythm and the activity rhythm may desynchronise. When this occurs cognitive performance may stay entrained to the sleep-wake cycle while the temperature rhythm runs-free, thus the rhythms have become 'fractionally desynchronised'. With fractional desynchronisation it is possible to determine the ranges of entrainment of different overt rhythms separately, independent of activity and of each other (Wever, 1979).

1.4.4 Partial Entrainment and Re-entrainment of Rhythms

It is also possible to exhibit partial entrainment where one rhythm remains entrained while another runs free. Aschoff and Wever (1981) explained that during re-entrainment of the rhythms, all circadian rhythms may move in the direction of the shift in the zeitgebers, however it is also possible that the system will split and run in opposite directions. Re-entrainment causes a transient state of internal temporal disorder in the subject. Because of this Aschoff and Wever (1981) believed that performance decrements are likely during this period. Although in contrast to this view, it has been shown that subjects actually perform better during the time that their rhythms are desynchronised and report a subjective feeling of contentment during this time (Wever, 1982).

1.5. Controlling Processes

According to Folkard (1996) the fact that the temperature rhythm and sleep wake cycle can run with different periods has led to the suggestion that human circadian rhythms comprises two or more processes:

1. A weak exogenous component, which is more prone to external influences. This is dominant in controlling the sleep wake cycle.
2. A stronger endogenous body clock. This is dominant in controlling the body temperature rhythm and is unaffected by external factors.

There appears to be general agreement that some circadian rhythms are controlled mainly by endogenous factors, while others are controlled mainly by external factors. For instance, heart rate seems to be almost entirely due to differences in activity level (it is exogenous in origin) while body temperature is at least partly controlled by the endogenous body clock. The endogenous clock exerts a much greater influence on the weaker exogenous process than vice versa (Folkard, 1996). Consideration will first be given to the exogenous components that influence rhythms, before moving on to consider endogenous components.

1.5.1 Exogenous Processes

The endogenous body clock and the exogenous processes are entrained to a 24-hour period by zeitgebers, such as the light dark cycle, knowledge of clock time and the behaviour of others. As a result, all our circadian rhythms usually show fixed phase relationships to each other (Folkard, 1996).

Wever (1979) found that some peoples' circadian rhythms could not be entrained by an artificial 24 hour light dark cycle, unless they were specifically asked to go to bed when the lights faded and to get up at dawn. This was interpreted as showing that although, for lower animals, the light-dark cycle is the most powerful zeitgeber (Folkard, 1996), for humans the best zeitgebers were of an informative or social nature (Wever, 1979).

1.5.2 Endogenous Processes

Oatley and Goodwin (1971) stated that an important function of rhythms in human performance is the synchronisation and timing of interrelated events. The brain must have a way of ensuring that events occur in the right order. Timing is required here and for timing oscillatory processes are needed. In support of this notion Oatley and Goodwin (1971) highlighted that the periodic patterns observed in EEG's, especially during sleep, indicates the existence of neuronal oscillations. Thus, the endogenous components that produce rhythms are termed oscillators.

In support of Oatley and Goodwin (1971), Wever (1979) believed both the temperature rhythm and the activity rhythm to be of oscillatory origin. Wever (1979) concluded, "human circadian rhythms are controlled by self-sustained oscillators" (p.43), but conceded that there is a co-ordination between the oscillators and chains of functions, for example body temperature can be considered an autonomous rhythm that can be separately controlled by a feedback mechanism of heat production and loss. More recently, Budzynski and Bingman (1999) found support for oscillatory control of circadian rhythms in pigeons.

Wever (1979; 1982) believed that the existence of internal desynchronisation demonstrates that all the circadian rhythms found in man are not controlled by just one oscillator, but rather that there are several oscillators and this gives them the ability to oscillate independently of each other causing internal desynchronisation. However, in order for the rhythms to synchronise Wever (1979; 1982) believed the oscillators are mutually interacting especially in the absence of zeitgebers, but the existence of internal desynchronisation shows that this interaction is only possible within certain limited period ranges.

1.6. Time of Day and Human Psychological Measures and Task Performance

We have discussed the entrainment of biological rhythms to those rhythms present in the environment, the possibility follows that the rhythms present within the brain, for example those shown by EEG's, may also become entrained to environmental rhythms. Evidence that brain rhythms can become entrained may come from the fact that lights flashing at a particular frequency can induce epileptic seizures. Such indications of the existence of entrainment in the brain have important implications in analysing time-of-day effects in human performance. Some work environments have machinery emitting sound and/or vibrations at certain frequencies that are suitable for encouraging sleep (Oately and Goodwin, 1971). Hence, it further follows that if rhythms intrinsic to the brain can be entrained by environmental factors, then they may also become entrained by biological rhythms, whether endogenously or exogenously controlled. Indeed, Oatley and Goodwin (1971) believed that daily rhythms of sleeping and waking and other associated biological rhythms are of importance in terms of human performance.

The efficiency with which we are able to perform particular tasks is in fact known to vary over the normal waking day, performance and psychological measures have been shown to reliably vary across the day as our rhythms fluctuate (Lenne et al, 1997). Folkard (1983) believed that the observed predictable performance trends over the day are thought to reflect the underlying 24-hour rhythms in most biological functions. In support of this, human performance efficiency has been found to have a tendency to parallel the peaks and troughs of the temperature rhythm, although no causal relationship is believed to exist (Campbell, 1992). Similarly, some psychological measures have also shown a relationship with temperature (Owens et al, 2000). Furthermore, Colquhoun (1971) stated that time-of-day effects can be observed in functions which involve levels of nervous activity from the most simple to the more complex. Knowledge of the relationship between time-of-day and task performance is essential if we are to aim to improve quality of life (Smith, 1992). Further, such variation in performance efficiency has obvious important implications (Owens et al, 2000) for safety and risk management.

1.6.1 Time-of-Day and Mood

Lenne et al (1997) claimed that subjective measures of alertness, motivation and sleepiness all show variations according to the time-of-day. For example, alertness is low at 0600 hours and peaks at 1400 hours then steadily declines to reach a minimum at 0200 hours. Motivation shows a similar trend while sleepiness shows the reverse. In support of this Dinges, Pack, Williams, Gillen, Powell, Ott, Aptowicz and Pack (1997) found significant time-of-day effects in subjective sleepiness ratings in sleep deprived subjects and Casagrande, Violani, Curcio and Bertini (1997) found subjective sleepiness, tiredness and energy levels on the visual analogue scale (VAS) to be influenced by the time-of-day. Further, Monk, Fookson, Moline and Pollak (1985) found diurnal variations in sleepy, happy and tired mood measures, better mood was found to occur four hours after waking and then fell over the day, but no diurnal trends were found for sad, calm and tense mood measures, while Thayer (1989) did find a diurnal rhythm in tension measures. Consequently, Owens et al (2000) concluded that although diurnal variations in alertness has been repeatedly demonstrated, whether any time-of-day effects exist in other mood measures is unclear. Owens et al (2000) pointed out that the majority of studies investigating the effect of time-of-day on mood and performance have used subjects who were living normally in their environment, therefore it's possible that variations between subjects in exogenous factors, such as meal times and sleep length and timing, could have influenced the obtained trends.

Owens et al (2000) stated that recent studies using chronobiological methodologies have found similarities in the diurnal variation in core body temperature and some psychological measures (for example, Johnson, Duffy, Dijk, Ronda, Dyal and Czeisler, 1992; Dijk, Duffy and Czeisler, 1992; Monk et al, 1997). However, they argued that although temperature and psychological measures may appear to be similar at a general level, a closer look at the trends frequently shows that the timing of the peaks and troughs in measures of performance differ from those seen in core body temperature. In support of this Monk, Buysse, Reynolds, Berga, Jarrett, Begley and Kupfer (1997) found that the sinusoidal (sine wave) patterns in core body temperature were not found in psychological measures. Further, Owens et al (2000) stated that subjective alertness measures show a sharp increase between 0800 and 1000 later reaching a peak at about midday before falling again during the rest of the day. It is stated that this peak in subjective alertness occurs several hours before the peak in body temperature. Furthermore, catecholamine levels also peak before body temperature.

As a result of such inconsistencies, Owens et al (2000) investigated diurnal variation in various psychological and performance functions and core body temperature. The subjects were highly practised women living under a controlled environment where the times of their sleep and meals were controlled. Ten of the participants experienced the natural light/dark cycle, while fourteen were exposed to no daylight but did have access to a clock. Various mood and performance tests were completed every two waking hours. The subjects were thus entrained to a 24-hour day without the presence of uncontrolled exogenous factors that might mask the true nature of the diurnal trends in the psychological measures. Owens et al (2000) found significant time-of-day variations for many of the measures with a post-lunch dip at 1600 for some variables. The nature of the time-of-day variation was different for different measures. The authors also found that many of the measures showed 'sleep inertia' or 'wake-up' effects and a 'sleep anticipation' effect, although the latter effect was smaller. The scores on the measures improved from waking at 0800 to 1000 hours and reduced from 2200 to 0000 hours (last measure). Owens et al (2000) suggested that the post-lunch dip in many of the measures of performance might be due to the controlled lunchtime. However, Wever (1979) stated that such rhythmicity was unlikely to be related to the ingestion of food, although some work has shown that meal content can influence the post-lunch dip (for example, Craig et al, 1981). It was found that all the variables showed different time-of-day trends to core body temperature. Although a similarity was found across subjects in the time-of-day trends in body temperature and the subjective ratings of alertness and cheerfulness obtained. Owens et al (2000) concluded that little can be inferred about time-of-day trends in many performance measures just from body temperature estimates, although some psychological measures may show similar trends to the temperature rhythm.

Thus, from such literature it seems clear that time-of-day variations can be observed in some psychological measures, for example alertness, although for measures such as tension, the evidence is more controversial. What seems to be becoming very clear is that the diurnal trend seen in body temperature cannot be relied upon to provide parallels to which psychological or performance fluctuations can be compared. As a result any study using temperature as a parallelism to any psychological or performance measure should be interpreted with care. Diurnal fluctuations in performance measures will now be considered in detail.

1.6.2 Time of Day and Task Performance

Guermess, Leconte and Leconte (1993) found evidence that participants showed greater efficiency for certain tasks, either in the morning or afternoon, concluding that these systematic fluctuations in performance efficiency mean that temporal factors cannot be ignored in the consideration of human performance. Indeed, some tasks are very sensitive to time-of-day effects, for instance Casagrande et al (1997) investigated the effectiveness of a letter cancellation task (LCT) in showing the effect of one night's sleep deprivation, effects of sleep-loss and time-of-day effects. It was found that the letter cancellation task was a sensitive measure of time-of-day effects and of one night's sleep deprivation. Visual analogue scale measures were also obtained but little correspondence was found between performance (LCT) and the subjective (VAS) measures. Similarly, Casagrande, Curcio, Tricarico, Ferrara, Porcu and Bertini (2000) found that all forms of the letter cancellation task showed time-of-day effects, but the 3 target letter format with a fixed time showed the most sensitivity to diurnal performance fluctuations.

Altabet (1995) stated that rhythmic variations in body temperature and an individual's preference for a particular time-of-day have been found to correlate with various aspects of performance. Similarly, Chelminski (2000) stated that research has suggested that some cognitive processes are subject to a circadian rhythm, with peak performance times correlating with physiological arousal patterns. Accordingly, Monk (1979) reviewed temporal effects in visual search and found a parallelism between the diurnal rhythm seen for simple repetitive task performance (for example, LCT) and that seen for the temperature rhythm.

Hockey and Colquhoun (1972) claimed that Kleitman (1939, 1963) conducted the first well-controlled experiments investigating time-of-day effects in task performance. Kleitman (1939, 1963) was also among the first to propose a relationship between performance and temperature. Kleitman (1939) stated, "the individual's capacity for doing mental or physical work is not the same throughout the waking period has been known for a long time" (p.150). Kleitman (1939) found an improvement over the morning with a fall in performance occurring during the afternoon. The peak in performance appeared to be at about midday. Temperature followed the trend seen in performance in the morning but became a little dissociated from it during the afternoon, peaking at about 1800 hours. He used simple perceptual-motor tasks such as letter copying and card-sorting. His subjects did the tests five times a day, but one subject who did the tests 10 times a day had a performance trend that followed his physiological trends quite well, with both peaks occurring at about midday. Temperature had a greater effect on choice reaction time, from which it was stated that this is "a definite suggestion of variation in mental work with body temperature" (p.160). Kleitman (1963) proposed two interpretations of this relationship. Either mental processes result from chemical reactions or the speed of thinking is dependent upon the metabolic activity of the cerebral cortex cells and that by increasing the latter by increasing body temperature, thought processes are indirectly quickened. Kleitman (1963) reviewed the literature and concluded "...there is a 24 hour variation in the state of the nervous system which determines the degree of mental fatigability at different hours of the waking period. Highest body temperature, best performance rate and now lowest fatigability are all reached almost synchronously in the afternoon. Both morning and night hours show the opposite characteristics, but in general the temperature is lower, performance is worse and fatigability is greater immediately upon getting up in the morning than they are at night just before going to bed... failure to sleep during the night will be reflected in performance during the following day" (p.156-157). Furthermore, Kleitman (1963) stated "speed and accuracy of performance may be slightly better or poorer on getting up in the morning than before going to bed at night, but they are low in either case, the peak or plateau appearing in between" (p.158). As discussed earlier, Kleitman (1939, 1963) believed the relationship between performance and temperature to be a causal one. Kleitman (1963) concluded "the ability to respond promptly appears to be best in the middle of the day, when the temperature is highest and poorest in the morning and late evening when body temperature is lowest." (p.154). Kleitman (1963) continued "a fairly good relationship between body temperature and reaction times was observed regardless of the time-of-day, indicating that there is probably no reaction time curve independent of the temperature" (p.154).

However, Hartley and Shirley (1976) found other simple repetitive tasks to show a 'U-shaped' function that did not parallel the diurnal rhythm in temperature, while, Klein, Wegman and Hunt (1972) found a continuous decline over the day in simple reaction time. Additionally, Hollingworth (1914) found performance in cognitive tests to show a decrease in efficiency from early morning to the end of wakefulness. Similarly, Englund (1979) found reading rate to be faster in the morning, decreasing over the day. Further, Gates (1916) found the efficiency curve of students improved up to midday, fell off after lunch, reached a maximum mid-afternoon and then showed a downward trend until the end of the school day, a pattern that again, does not parallel the diurnal trend seen for temperature.

Not surprisingly, contemporary thinking on the relationship between performance and temperature is a little more advanced. Clearly, there are inconsistencies with some research suggesting that performance rhythms parallel the temperature rhythm while other research suggests that this is not the case. It seems that it is the nature of the task that is of more importance in determining the exact nature of the performance curve. Procedures used in early experiments, that frequently found similarities between performance and temperature trends, often used boring vigilance tasks requiring no higher level processing. As factors such as memory load and type (short- or long-term, see section 1.7) became involved, the exact course of the performance efficiency curve across the day varied (Campbell, 1992). In some cases, for instance those tasks requiring short-term memory processes, the relationship between core body temperature and performance efficiency is an inverse one (Folkard and Monk, 1980; Laird, 1925). Additionally, the type of measure used to evaluate the results is of great importance, for instance Monk and Leng (1982) found that if reaction time is used as a measure of performance then the usual relationship with temperature was seen where participants became faster over the day, yet if consideration was given to accuracy rates then the opposite relationship was seen. Consistently, Folkard (1975) suggested that measurements of speed and accuracy may represent different components of diurnal performance both with a different relationship to arousal level.

Establishing exactly how demands of a task influence the diurnal performance trends seen and establishing if physiological trends parallel these patterns has important implications for predicting the likelihood of, and preventing, accidents occurring during completion of tasks performed during the course of everyday life, especially those done at all times of the day (and night), for example driving. Schwing (1990) stated that circadian rhythms are one possible contributor to road accidents and McDonald (1989) stated that the time-of-day is one variable that shows a strong association with accident rates.

Lenne et al (1997) stated that reviews of data on driving accidents report a prominent time-of-day effect, with the greatest number of accidents occurring between 0300 and 0500 hours, and with a smaller secondary peak occurring between 1400 and 1500 hours. The first peak corresponds to the nadir in body temperature, supporting the notion that performance on some tasks parallels daily temperature fluctuations, while the secondary peak corresponds to the post lunch dip. Folkard (1997) confirmed the existence of a circadian rhythm in the risk of road accidents finding, in support of Lenne et al (1997), a major peak in accidents at 0300 hours. Folkard (1997) believed this could not be attributed solely to drivers falling asleep at the wheel as wakeful subjects showed a similar circadian dip at this time. Thus he concluded that the 0300 peak in road accidents reflects lowered capabilities. However he points out that there are secondary peaks in accident rates that are difficult to account for in terms of circadian rhythmicity and suggests that a time on task effect may be responsible for these where risk of an accident increases with increased time spent on the task.

Lenne et al (1997) found reaction time to be affected by time-of-day, showing an impairment in performance at 0200 and 0600 hours with improvements in between 1000 and 2200 hours and an early afternoon dip, thus these results support the notion that driving performance is subject to diurnal variations. This was supported by Jaencke, Musial, Vogt and Kalveram (1994) who found a deterioration in driving performance early in the day. In a later study, Lenne, Triggs and Redman (1998) found that performance decrements seen in driving ability were the outcome "of an interaction between sleep deprivation and the time-of-day". They found that performance improved steadily across the day between 0800 and 2000 hours and this rhythm in performance over the day was similar after normal sleep and sleep deprivation. Interestingly, Summala and Mikkola (1994) found the peak time for accidents to vary according to the age of the driver. The peak time for accidents in drivers 18 to 20 years old fell between midnight and 0600, while for those who were 56 years or older the peak time occurred during the late afternoon. Thus the peak time for accidents among younger drivers coincides with the nadir in body temperature and supports Lenne et al's (1997) and Jaencke, Musial, Vogt and Kalveram (1994) findings, while the peak time for accidents among older drivers may correspond with the post lunch dip that can persist until 1600. In contrast to these findings, Horne and Baumber (1991) failed to uncover a time-of-day affect in driving ability, finding that the time-of-day did not affect lateral corrective steering movements. A probable reason for this inconsistency may involve the different aspects of driving and therefore the different task demands, that the authors chose to monitor (reaction time versus corrective steering movements).

An important point regarding research into time-of-day effects is highlighted by Owens et al (2000) who believed that diurnal variations in performance (and mood) measures may be masked by environmental and behavioural factors. Measures of performance show distinct practice effects that could mask the nature of any existing time-of-day effect. Although certain methods can be adopted to help balance practice effects, these methods make the assumption that different subjects show similar practice and time-of-day trends. These methods also assume that the time-of-day effect found in unpractised subjects could generalise to situations in real life where people carry out highly practised tasks such as driving.

1.7. Time of Day and Memory

A criticism of Kleitman (1939,1963) has been that he based his conclusions primarily on simple perceptual-motor tasks not more cognitively orientated ones (Folkard, 1983). Performance of other, non-perceptual motor tasks, have also been shown to exhibit diurnal variation. Much recent research has focused on time-of-day effects in the performance of memory tasks (Owens et al, 2000).

There is evidence to suggest that the time-of-day affects our ability to remember, for example Hockey and Colquhoun (1972) stated that memory appears to be better in the morning when temperature is at a low point. Also, Poirel and Larouche (1987) believed that the processes of forgetting are dependent upon the circadian stage. However as we shall see below, this effect seems to be different for short- and long-term memory (Berger, 2000). It is important to note that although researchers frequently differentiate between different types of memory, this distinction can sometimes be unclear, for example working memory often involves a combination of not only immediate processing and short-term memory but some working memory tasks may also involve semantic memory, for instance.

1.7.1 Time of Day and Immediate/Short-Term Memory

Studies involving immediate/short-term memory have involved participants reading short prose and then immediately being given a test of memory for the contents of the prose (Folkard, 1983). Berger (2000) stated that the early morning tends to be associated with better immediate memory, which is associated with the low arousal levels connected with this time-of-day. Indeed, several early studies have found memory to be better in the morning. Winch (1912) found a consistent but small advantage in the morning for immediate memory compared to the afternoon. Similarly, Blake (1967b) and Baddeley, Hatter, Scott and Snashnall (1970) found adult performance on an immediate memory task was better in the morning than the afternoon, supporting the findings of Ebbinghaus (1885) who found learning serial lists of nonsense syllables consistently tended to be more rapid in the morning. Likewise, Folkard, Monk, Bradbury and Rosenthal (1977) found that recall of information in the short term is also better in the morning in children. Blake (1967b) also found performance on a digit-span task to decrease over the day. Baddeley et al (1970) summed up early research by concluding that the efficiency of immediate memory varied as a function of time-of-day and suggested that this effect was due to circadian fluctuation in the level of arousal with the efficiency of short-term memory being impaired when arousal is high. Finally, Monk and Folkard (1978) found immediate memory for a film to be better at 0400 (when arousal is low) than at 2030 (when arousal is higher), however individuals who watched a film at 0400, forgot more details over a 28-day period (this is consistent with literature suggesting that low arousal is associated with poorer delayed memory and supports the notion that the amount of information that can be remembered is dependent on the time-of-day at which the material is learned rather than at which time-of-day it is tested, see section 1.7.2). Later studies suggest a more complex picture. Natale and Lorenzetti (1997) found that immediate memory was superior in the morning for morning types and superior in the afternoon for evening types, suggesting an interaction with individual differences is possible.

As noted earlier, Folkard, Knauth, Monk and Rutenfranz (1976) stated that typically, early studies have found that body temperature and performance efficiency reveal very similar circadian patterns (Kleitman and Jackson, 1950; Colquhoun, Blake and Edwards, 1968a, 1968b), but these tasks have placed little reliance on short-term memory. It is stated that while performance on most tasks improves over the course of the waking day (for example, Blake, 1967b) performance on short-term memory tasks decreases. Folkard and Monk (1983) believed that studies that have found performance has *improved* over the day (for example, Lenne, 1997) have not used specific memory tasks, thus accounting for this inconsistency. Folkard et al (1976) believed that the reason why short-term memory tasks have such a different effect on the circadian variation in performance efficiency is concerned with arousal. Folkard et al (1976) believed that it could be concluded from Blake's (1971) studies that arousal increases over the day and it is this that mediates the increase in performance efficiency over the day seen for simple immediate *processing* tasks such as visual search. So studies using immediate *processing* show performance to be at its' best at the end of the day when arousal is high. Conversely, Folkard et al (1976) stated that a number of studies have found immediate/short-term *memory* to be impaired under high arousal (Mclean, 1969; Walker and Tarte, 1963). Folkard et al (1976) continued, that if arousal increases over the day, this could account for the concurrent decrease in short-term memory performance. Further, if changes in temperature correspond to changes in arousal, then a negative correlation should exist between temperature and short-term memory (Folkard et al, 1976).

Furthermore, research suggests that the *memory load* (that is, the amount of information that has to be remembered to complete the task) of a task may play an important role in the relationship between temperature and performance. Folkard et al (1976) found that performance was highly positively correlated with body temperature when memory load was low, but when memory load was high performance negatively correlated with temperature. With medium memory load there was no relationship between performance and temperature. Thus the memory load of the task needs to be taken into account.

1.7.2 Time of Day and Delayed/Long-Term Memory

While there is some evidence that short-term memory may *deteriorate* under high arousal, it has also been found that delayed memory (over 15-20 minutes) may be *superior* when the material is learned under high arousal. This suggests that while immediate memory is better in the morning, long-term memory should be better when the material to be remembered is presented in the afternoon/evening (Folkard, 1983). Similarly, Baddeley et al (1970) stated that a general tendency could be found for the retention of items processed under high arousal to be better after a long delay than after a short one. An important point to note here is the difference between 'time of presentation' and 'time of retrieval' (when discussing long-term memory this distinction is possible whereas in the case of immediate memory this distinction cannot be made as presentation and retrieval obviously occur at the same time-of-day), as just discussed it seems retention is better when presentation has taken place later in the day under higher arousal levels. It seems that time of retrieval, whilst exerting some effect, has much smaller effects than those of time of presentation. Clearly then, time of presentation is of more importance than time of retrieval when considering time-of-day effects in delayed retention (Folkard, 1983).

For example, Folkard and Monk (1980) found evidence that long-term retention is better if the material is given in the afternoon, when arousal is higher, regardless of the time-of-day of recall. In support of this, Berger (2000) stated that the late afternoon and early evening tend to be associated with better long-term memory, which is related to the high arousal levels associated with this time-of-day. Folkard (1983) suggested that a possible explanation of these time of presentation effects is that the superiority of immediate memory during the morning is due to more attention being given to the physical characteristics (when arousal is lower) of the information and the superiority of delayed memory when the information to be remembered is presented later in the day is due to more attention being given to the meaning of the material (when arousal is higher). It is highlighted that this would account for the failure of Ebbinghaus to find a time-of-day effect in delayed retention (as he used nonsense syllables that do not convey meaning) and would also account for larger time-of-day effects being observed with more realistic and therefore more meaningful material.

Yet, in contrast to previous findings, Gunter et al (1984) claimed that time of presentation of material has only a slight effect on performance levels. Similarly, Millar, Styles and Wastell (1980) found support for the time of *retrieval* being more important than the time of *presentation.* While, Holloway (1967) failed to find an advantage of time of presentation upon students' recall of information in a multiple-choice test.

The everyday applicability of such research is highlighted by Berger (2000), who referred to individual differences in arousal patterns, for instance the morningness-eveningness difference. As adolescents are often evening types, their peak arousal periods occur around the late afternoon-early evening, resulting in the majority of their scholastic day occurring during periods of low arousal. Berger (2000) highlighted the fact that many adolescents attending schools with early starting times seem to be chronically sleep deprived and that it has been suggested that a change in arousal has a greater effect on performance when the overall arousal level has been decreased due to sleep deprivation. It therefore follows that adolescents attending early start schools may be especially susceptible to time-of-day effects on memory. In contrast to previous research, however, Berger (2000) found no support for the notion of a time-of-day effect on short or long-term memory, further, memory performance was not related to the number of hours sleep taken on the previous night nor was memory performance related to subjective classification of morningness-eveningness. Nonetheless, subjective tiredness perceptions corresponded to the earliest class times in the morning and to the times following lunch.

1.7.3 Time of Day and Working Memory

Working memory tasks involve both immediate processing and short-term memory (Folkard, 1975, 1983). Verbal reasoning and mental arithmetic tasks are the most commonly used in time-of-day research into working memory (Smith, 1992), presumably because such tasks require both immediate processing and short-term memory (Folkard, 1983). It has been found that performance on these tasks peaks at about midday (Laird, 1925; Owens et al, 2000; Folkard, 1975), while performance on simple serial search tasks, such as proof reading, that involves very little or no memory peaks in the evening (Folkard and Hill, 2002).

It should be noted here that the time-of-day trend seen in immediate *memory* (for example, where an immediate test of memory is given for the contents of an article just read) is practically the opposite of that seen in immediate *processing* (for example, the LCT) tasks (Folkard, 1983). Folkard (1983) claimed that the trend in performance for working memory tasks falls between the decreasing time-of-day function shown by short-term/immediate *memory* and the increasing time-of-day function shown by simple immediate *processing* tasks. However, Folkard et al (1976) and Folkard (1983) found that the peak in working memory performance occurs earlier on more highly loaded memory tasks. In support of this, Davies, Parasuraman, Toh (1984) found performance on a memory-loaded task to be better in the morning than in the afternoon. In addition to this Folkard (1983) pointed out that the memory load involved in a task would probably be affected by personal characteristics such as practice, age and intelligence. Smith (1992) agreed that studies have also shown that performance on working memory tasks show different trends according to the level of practice or ability of the subject. The morningness of the subject also interferes with peak performance trends. For example Monk and Leng (1986) found that performance on a logical reasoning task (a task requiring working memory), revealed an interaction between time-of-day and morningness, with morning types showing a peak in performance at 0800 and evening types showing a later peak. Therefore, once again when investigating diurnal variations in performance on working memory tasks the demands of the task must be considered alongside the personal characteristics of the subjects (Owens et al, 2000). Furthermore, Folkard and Hill (2002) stated that working memory tasks use several different cognitive sub-systems such as short term information storage, information processing and throughput and as such it is plausible that the diurnal pattern observed is the "outcome of a combination of different trends associated with the different cognitive mechanisms involved" (p.57).

1.7.4 Time-of-Day and Semantic Memory

Semantic memory refers to memory for meanings, it is often associated with long-term memory because memories in the long-term store are often remembered on the basis of their meanings (Folkard, 1983). It has been found that when recognising dominant and non-dominant instances of categories, responses are faster later on in the day, with the greatest difference between morning and evening being found with low-dominance items (Millar et al, 1980; Tilley and Warren, 1983). Smith (1987a) employed a category instance retrieval task using dominant and non-dominant instances and found performance to be quicker later in the day, although no interaction was observed between dominance and time-of-day. Furthermore, other studies using semantic processing tasks have also found semantic memory to be quicker later in the day (Smith, 1989). Interestingly, Smith (1992) noted that the circadian variation seen for semantic memory is similar to that seen for perceptual-motor tasks. Additionally, long-term memory performance has also been shown to be better later in the day and this is consistent with the notion of semantic memory being synonymous with long-term memory (Smith, 1992).

There is evidence that the time-of-day effects seen in semantic memory tasks can be altered by changing the nature of the task. For instance, Smith (1987a) found performance on a category instance task was faster later in the day when participants were given separate lists of dominant and non-dominant instances. If these were included in the same list so the participants had to constantly use different retrieval strategies, then the time-of-day effect disappeared. Similarly, Whitney and Williams (unpublished) examined whether semantic retrieval varied with the time-of-day when participants were required to make same-different decisions for category-exemplar pairs. Circadian variations were not observed for different category pairs, this was attributed to different decision processes being used for different judgements. However, same category judgements were most efficient in the evening consistent with previous findings and the arousal theory. Slowest semantic access was mid-afternoon, possibly resulting from a post-lunch dip. Whitney and Williams (unpublished) concluded that the tendency for the level of processing used to shift with the time-of-day could be related to circadian patterns in the accessibility of semantic information and that the effect of the time-of-day is not the same for same and different judgements.

In sum, immediate/short-term memory tasks reveal better performance in the morning while arousal is low (Ebbinghaus, 1885; Winch, 1912; Baddeley et al, 1970; Monk et al, 1978), as arousal increases performance on these tasks *deteriorates* (Folkard, 1983). Conversely, performance on delayed/long-term (15-20 minutes) memory tasks is *superior* when material is *presented* under high arousal (Baddeley et al, 1970; Folkard and Monk, 1980; Folkard, 1983; Berger, 2000). Performance on working memory tasks peaks at midday (Laird, 1925; Owens et al, 2000; Folkard, 1975) but the peak in performance occurs earlier as the memory load involved in the task increases (Folkard et al, 1976; Folkard, 1983). Finally, performance on semantic memory tasks is quicker later in the day (Millar et al, 1980; Tilley and Warren, 1983; Smith, 1987a, 1989), however the observed time-of-day effect can be altered by changing the nature of the task (Smith, 1987a).

1.7.5 Changing Strategy

Time-of-day effects can also depend on the way the task is performed. For instance, time-of-day effects may not be a result of passive changes occurring because of processing limitations, but may be due to different strategies being employed at different times of the day. It is currently unclear as to whether strategy changes occur because of the influence of endogenous rhythms or whether changing strategy is an attempt to maintain a good level of performance in a below optimal state (Smith, 1992).

Evidence that people change their strategies at different times of the day came from research showing that individuals become faster but less accurate later in the day (Monk and Leng, 1982; Blake, 1971; Craig and Condon, 1984, 1985). To examine whether changing strategy is under voluntary control or whether it reflects a passive response, Smith (1991) attempted to alter the speed-accuracy trade off sometimes observed in time-of-day studies by giving participants different instructions. Participants were either told to respond as quickly and as accurately as possible or speed or accuracy was stressed individually. He found that performance was faster but less accurate in the early evening over all conditions showing that the speed-accuracy trade off was not a result of subjects choosing to use certain strategies at certain times. However, Folkard (1979) showed that if subjects are instructed to use a certain strategy to complete immediate recall tasks, the time-of-day effect was eliminated (Folkard, 1979).

Furthermore, there is evidence to suggest that the changes seen in short-term memory and long-term retention over the day reflects a change in the level of processing that the individual spontaneously engages in (Folkard, 1982). For instance, the finding that time-of-day can influence ability to access semantic information is compatible with results from episodic memory (memories that are stored along with information about where and how they were formed) studies where more reliance may be placed on acoustically based maintenance processing during the morning (when arousal is lower) and on semantically based elaborative processing in the evening (when arousal is higher) (for example, Folkard, 1979). Lorenzetti and Natale (1996) provided further support, finding that elaborative and integrative processes are used more in the afternoon and maintenance processes of superficial linguistic form is employed more in the morning. Further, Folkard (1980), Oakhill (1986a, 1986b, 1988) and Marks and Folkard (1988) have all provided evidence that more 'maintenance processing' is engaged in during the morning while more 'elaborative encoding' is used during the afternoon.

Finally, it has been suggested that changes in the performance of perceptual motor tasks also reflect changes in the strategy used to complete the task over the day. Monk and Leng (1982) attempted to differentiate between two explanations of results, one was the 'capacity' explanation that proposed that changes occurred in the *rate* of information processing over the day, and the second was the 'strategy' explanation that proposed that changes occurred in the *amount* of information processed at each decision point. Monk and Leng (1982) believed the strategy explanation to be more suitable to account for temporal changes in simple repetitive task performance because this accounted for the variety of time-of-day functions that can be observed for simple repetitive tasks and allows for the exact time-of-day function to depend on the strategy change that occurred.

Consequently, all this suggests that time-of-day effects in performance may result from changes in the way tasks are conducted rather than changes in the efficiency of the process used (Smith, 1992).

1.8. The Arousal Theory of Time of Day Effects in Performance

As noted earlier, the best time-of-day to perform a task seems to depend, among other things, on the memory load involved in the task (Monk, 1982); a low memory load task (for example, a deletion task [Blake, 1967b]) will reveal an evening peak, while a high memory load task (for example, remembering information in a prose for a short period [Folkard and Monk, 1980]) will show a morning peak in performance. Monk (1982) stated that the accepted explanation of such fluctuations in performance is the arousal theory.

Kleitman (1939; 1963) showed temperature to parallel performance on simple tasks, except for the post-lunch dip. This gave rise to the arousal model, for which Colquhoun (1971) put forward the best case (Folkard, 1983). Upon finding that people often reported subjective feelings of sleepiness at the time of the post-lunch dip, Colquhoun (1971) equated sleepiness with arousal and proposed that arousal was the underlying factor that mediated the relationship between the temperature and performance rhythms. Both body temperature and arousal increase from a minimum at about 0500 to reach a maximum at about 2000, therefore Colquhoun (1971) asserted that there was a circadian rhythm in arousal that showed a rise over the course of the waking day with an early evening peak, thus (with the exception of the post-lunch dip in arousal) paralleling the circadian rhythm in body temperature. Consequently, it was postulated that arousal could account for the post-lunch dip seen in performance that is not reflected in the temperature rhythm. Thus although Colquhoun (1971) rejected the notion of a causal relationship between temperature and performance he believed that circadian fluctuations in body temperature paralleled those in performance with the exception of the post lunch dip. Colquhoun's

(1971) model was based on findings of a relationship between temperature and performance on simple tasks and it was concluded "the form (or even the existence) of rhythms in cognitive or 'higher-level' functions such as memory load....still requires to be established" (p.100).

Subsequently, rhythms in the performance of these "higher-level functions" have been interpreted as reflecting the underlying circadian rhythm in arousal. The Yerkes-Dodson Law (1908) postulated that arousal and performance can be related by an 'inverted U' and argued that the optimum level of arousal at which to perform a task was dependent upon the complexity of the task. The optimum level of arousal was therefore high for low memory load tasks and low for high memory load tasks. Thus as Colquhoun (1971) proposed that arousal increased over the day from a minimum in the morning to a maximum in the evening, then it follows that complex/high memory load tasks are best performed in the morning while simpler/low memory load tasks are best performed in the evening. Monk (1982) stated that the parallelism between the arousal and temperature rhythms allows arousal to be predicted from temperature, and that the change in the optimum level of arousal with high memory load can be seen as a change in this relationship between temperature and performance rhythms. This change has been demonstrated by Folkard et al (1976). Thus assuming temperature and performance rhythms are related, being mediated by the arousal rhythm, then it follows that if a high level of arousal is best for low memory load tasks then these tasks will be performed best at high temperatures (that is, later in the day). Similarly, if a low level of arousal is best for high memory load tasks then these tasks are performed best at low temperatures (that is, earlier in the day). Meanwhile, intermediate memory load tasks will show no relationship between temperature and performance.

In sum, the arousal theory framework holds that the more complex a task is (and therefore the higher the memory load), the earlier in the day performance should peak, while arousal is low. Conversely, the simpler a task is (and therefore the lower the memory load) the later in the day performance should peak, while arousal is high. Simple task performance should be positively correlated with temperature, more complex task performance should be negatively correlated and intermediate task performance should show no relationship with temperature. Thus the arousal theory can account for many different trends in performance, depending on complexity (or memory load), over the day.

Colquhoun's (1971) unidimensional theory used the same theoretical angle as that used in interpreting the effect of other factors hypothesized to increase (for example, noise) or decrease (for example, sleep deprivation) arousal. Therefore, Colquhoun's (1971) theory has generated many predictions that are easily tested and can also account for a variety of results, consequently however, a problem with the inverted U theory's ability to predict performance is that the theory is difficult to falsify (Folkard, 1983). The unidimensional theory has received much empirical support, however it has also accumulated much criticism (see Folkard and Monk, 1981).

Evidence in support of the theory has found the time-of-day to interact with other factors that were assumed to either increase or decrease arousal level (Folkard, 1983), for example, sleep-deprivation. The inverted U predicted that as arousal increased so performance would improve up to an optimum level after which performance efficiency would decrease, thus presumably, after sleep deprivation arousal is lower thus accounting for the more pronounced time-of-day effects in sleep deprived subjects (Colquhoun, 1971). Further, Blake (1971, 1976) found support for the arousal model finding again that time-of-day effects in performance interact with other factors that influence arousal, for example extraversion and knowledge of results, in a way that can be accounted for by the inverted U and are thus consistent with the arousal theory. Colquhoun (1960) found similar evidence in support of the unidimensional model finding that introverts' detection rates on a visual vigilance task was better than that of extraverts in the morning, but in the afternoon the relationship was reversed. However, Revelle, Humphreys, Simon and Gilland (1980) found the interaction between time-of-day and extraversion to be confined to the impulsivity component, proposing that high and low impulsives may be phased differently in respect of the their time-of-day. Eysenck and Folkard (1980) actually found support for this, finding low impulsives' body temperature to peak earlier in the day. Nonetheless, further support was provided for the unidimensional model in that a personality variable that was known to influence arousal was found to interact with the time-of-day. Furthermore, Revelle et al (1980) found that caffeine (an arouser) had unfavourable effects on low impulsives during the morning, but had positive effects during the afternoon, while high impulsives benefited from caffeine during the morning but were hindered by it during the afternoon. This is again explained by the unidimensional theory, assuming that high and low impulsives follow different diurnal rhythms with a later peak occurring for high impulsives.

Similarly, the application of a stressor such as noise can influence performance in a way that is consistent with the arousal model (Coyle, 1989). Coyle (1989) stated that if a person is on the ascending part of the inverted U a stressor will move them towards the point for optimum performance, here the stressor increases arousal and improved performance. Conversely, if a person is on the descending part of the inverted U, a stressor will move them further away from the optimum level of arousal needed for peak performance, thus the application of the stressor decreases performance. Thus theoretically, a stressor/arouser should interact with time-of-day to produce better performance in the morning than in the evening on simple tasks. In support of this Blake (1971) found loud noise improved performance at 0800 but not at 1030.

Thus, as Smith (1992) stated, the arousal model appears to account well for interactions between time-of-day and other factors influencing arousal. Indeed, further support was provided for the unidimensional model by Blake (1967a) who examined the effects of the time-of-day on several performance measures (five-choice serial reaction, auditory vigilance, simple arithmetic, letter cancellation and card sorting). Performance on these tasks increased over the day peaking at the last testing time. Body temperature also showed this trend peaking at about 2000/2100, thus support was found for a central point of the arousal theory, that performance and temperature are related. Nevertheless, some studies have indicated that the unidimensional arousal model is not adequate (Smith, 1987a; Smith and Miles, 1985, 1986a, 1987a). Moreover, Coyle (1989) concluded that many of the studies that have been reported in favour of the unidimensional theory "did not allow equally for verification and falsification" (p.25).

Evidence that the time-of-day does not always interact with other factors influencing arousal concerns selectivity. Easterbrook (1959) suggested that as arousal increases, an individual decreases their level of cue utilisation so that irrelevant cues are ignored and performance improves. However, after an optimal level of arousal is reached, the individual begins to ignore relevant cues and so performance on the task declines. Easterbrook's (1959) theory suggested that selectivity would be greater later in the day. The Stroop colour word test, where in one form of this test, the colour of the ink in which an irrelevant colour word is written must be named, has been used to test this. Contrary to what the arousal theory would predict, it has been found that interference in the Stroop task increases over the day to peak at 2000 hours (Hartley and Shirley, 1976). No difference has been found in the interference task when late morning and early afternoon have been compared (Smith and Miles, 1987b) or when early evening and early morning have been compared (Smith, 1992).

Further evidence that fails to support the arousal theory is concerned with the assertion that body temperature is indicative of arousal. Firstly, Blake (1971) and Horne and Ostberg (1976) found small differences in the phase of the body temperature rhythms of introverts and extraverts and morning and evening types. Consequently, one might expect only a small difference in the performance of these groups, but a greater difference than expected was found. Eysenck and Folkard (1980) also found only a small difference in the phase of high and low impulsives body temperatures and this disagrees with Revelle et al's (1980) prediction of a difference in performance of many hours. Thus it seems temperature is an unreliable predictor of circadian performance due to the influence of personality variables (Coyle, 1989; the relationship between extraversion, morningness, circadian arousal patterns and performance is considered in greater detail in section 1.8.1.). Secondly, the temperature rhythm fails to mirror the drop in performance during the post-lunch dip and many studies have failed to find a parallelism between temperature and performance (for example, Folkard, 1996; Campbell, 1992; Folkard et al, 1976).

Other evidence against the arousal model concerns its' assertion that arousal peaks in the evening. Research has found arousal to increase during the morning peaking between 1100 and 1400 hours (Thayer, 1967, 1978; Clements, Hafer and Vermillion, 1976; Folkard et al, 1976; Akerstedt, 1977; Folkard and Monk, 1978) this is at odds with Colquhoun's (1971) original assertion that arousal peaks at about 2000 hours. Furthermore, physiological indices of arousal are indicative of a peak in arousal at midday. For example, Klein, Herrmann, Kuklinshi and Wegmann (1977) and Akerstedt (1978) found adrenaline to peak at 1200 hours, while Akerstedt (1978) also found noradrenaline to peak at 1200 hours.

In view of such difficulties as those outlined above, perhaps it is not surprising that Folkard (1983) believed the arousal theory to be over-simplistic. It seems the notion that daily variations in performance reflect the circadian rhythm seen in basal arousal only holds for a small range of simple tasks (Owens et al, 2000). The arousal theory of time-of-day effects has therefore been mainly rejected as a full and complete explanation of the phenomenon (Smith, 1992). Smith (1992) concluded that even when time-of-day effects are explained in terms of endogenous rhythms there is agreement that several rhythms or effects can influence performance. For example, alertness can be influenced by the amount of time elapsed since waking, as well as the sleep-wake rhythms. Also, exogenous factors must be considered, for example, a build-up of fatigue can influence performance later in the day.

1.8.1 Consideration of Other Factors

So we have seen that other factors that are assumed to influence arousal such as morningness or personality, have provided support for the arousal model. Although the arousal model has been mainly rejected as a complete explanation for time-of-day effects, the possibility still stands that it does exert at least some influence. Consequently, three factors that are assumed to influence arousal and affect performance will be considered in more detail here, before considering more contemporary theories of diurnal performance fluctuations. These are morningness and the influence of age, and personality.

As Folkard (1983) stated, it is well established that some people prefer to work in the morning ('morning types') while some prefer to work in the evening ('evening types'). Monk (1990) stated that two of the most important factors in determining when a person is likely to perform at their best are age and where they fall on the 'morningness-eveningness' scale. Predictably, extreme morning types have been shown to have greater difficulty in adapting to night work, while extreme evening types do not perform well early in the morning. Similarly, the performance levels of extreme morning types deteriorates over the day while the performance levels of evening types improves (Folkard and Hill, 2002). Age can play an important role, as people tend to show more morningness as they become older (Monk, 1990; Yoon, May and Hasher, 2000). May, Hasher and Stoltzfus (1993) for example, found that most of their younger participants were evening types while most of the older participants were morning types. During late afternoon younger but not older participants performed optimally but no age differences in memory performance were observed for the morning. Yoon et al (2000) indicated that if studying circadian rhythms in performance, the effect of age is an important variable, as not only do cognitive changes occur with age, such as decreased inhibitory functioning (the ability to ignore irrelevant information), but also circadian patterns themselves change.

Hasher, Zacks and May (1999) investigated the combined effects of inhibitory control, circadian arousal and age finding that for both young and old adults reduced inhibitory control was associated with performing tasks requiring inhibitory control at the persons' non-optimal time-of-day. Winocur and Hasher (1999) investigated ageing and time-of-day effects on cognition in rats, results suggested that circadian disruption resulting from old age can have an adverse effect on memory and cognitive functions related to it, which can influence inhibitory control. Chelmski (2000) proposed that the influence of time-of-day and age on inhibitory control was concerned with attentional deficits.

More recently, Waterhouse, Weinert, Minors, Folkard, Owens, Atkinson, Nevill and Reilly (2000) found a significant correlation between the score of an individual on a morningness questionnaire and the phase of the circadian rhythm of temperature, with the phases becoming earlier as the degree of morningness increased. In support of this, Kleitman (1939, 1963) divided people into two groups, one group that showed an early temperature peak (and presumably were morning types) and one that showed a later temperature peak (and presumably were evening types), and found differences in the performance rhythms that paralleled their temperature rhythms. However, this rests on the assumption that temperature reflects the level of arousal of the nervous system and that performance efficiency is related to an individuals' level of arousal (Blake and Corcoran, 1972). Further, again assuming that temperature reflects arousal level, Petros, Beckwith and Anderson (1990) found recall to decrease across the day for morning types but to increase across the day for evening types, this is consistent with arousal levels being highest in the morning for morning types and decreasing over the day, while arousal levels are assumed to be lower in the morning for evening types that increase over the day. Further, Altabet (1995) found temperature difference between test sessions correlated with time-of-day preference. In the morning participant's with a morning preference had higher temperature values than those with an evening preference and vice versa for the afternoon. Further, temperature difference related to verbal IQ, verbal comprehension and processing speed performance in a way that was similar to time-of-day preference. In support of the difference in temperature rhythms between morning and evening types, Corbera and Grau (1993) also found oral temperature curves to have larger amplitudes and to be phase delayed in evening-types. Conversely, Folkard (1983) stated that it seems that people do not actually differ that much in the timing of their temperature peaks, this led to the belief that only small differences would therefore be found between the performance rhythms for morning and evening types. Yet Horne, Brass and Pettit (1980) found that although extreme morning and evening types had similar temperature rhythms their performance rhythms were very different. Evening types showed a *positive* relationship between temperature and performance while morning types showed a *negative* relationship between temperature and performance.

As stated previously, where a person falls on the introversion-extraversion scale is another factor that is assumed to influence arousal level. Blake and Corcoran (1972) stated that most work suggests that there is some difference in the arousal mechanism in introverts and extraverts. The indication is that introverts have higher arousal levels than extraverts regardless of the time of day, however the work suggests that circadian rhythms are definitely involved. Blake (1967b) found the body temperature of introverts to be higher than that of extraverts during the morning and early hours of the afternoon, but the temperature of extraverts was higher than that of introverts during the evening. Gunter et al (1984) suggested that if introverts and extraverts have different levels of arousal then they might also show parallel differences in memory performance over the course of the day. Blake and Corcoran (1972) stated that the introversion-extraversion scale has been found to discriminate morning people from evening people. Blake (1976) aimed to determine whether these differences were associated with differences in their physiological circadian rhythm as indexed by body temperature. Blake (1976) found a relatively small relationship between personality, body temperature and the time-of-day, thus it was concluded that this might help to explain the performance differences observed between introverts and extraverts. For instance Colquhoun (1960) and Colquhoun and Corcoran (1964) found that introverted individuals performed better than extraverted individuals in the early morning, while the extraverts performed better in the afternoon while the introverts performed less well at this time. Blake (1971) confirmed these findings. However, it seems that this relationship between introversion-extraversion, morningness-eveningness and time-of-day is influenced by other factors, for example Colquhoun and Corcoran (1964) found that introverts were only better in the morning when the participants were tested in isolation while the presence of other subjects improved the morning performance of extraverts. These studies used simple performance tests, when a cognitive task is used, it has been found that it is the impulsivity component of the extraversion scale that produces the observed time-of-day effects (Revelle et al, 1980). Lawrence and Stanford (1999) highlighted the fact that previous research had found indications that high impulsive people (that is, extraverts) show better performance in the evening than in the morning because of different variations in their diurnal arousal rhythms, but this study failed to find significant interactions between the level of impulsivity of a person and the time-of-day on performance levels. In spite of this, in comparison of high impulsives and low impulsives, it was observed that high impulsives showed a greater variability in their performance and a faster 'cognitive tempo'.

Thus it seems that the time-of-day interacts with individual differences and therefore affects performance at different times of the day despite only small differences in the temperature rhythm (Folkard, 1983). This illustrates the number of additional factors that must be considered when interpreting time-of-day and performance literature and lends support to the notion that, even though research on personality and time-of-day effects are based on relative differences not absolute ones, the demands of the task and personal characteristics should also be considered when interpreting such studies (Owens et al, 2000). Additionally, many other factors that influence arousal level are also thought to interact with the observed trend in performance over the day. These include sleep deprivation, stress, noise, social isolation, knowledge of results, motivation, test expectation and drugs (Folkard, 1983; Hancock, 1989; Breen-Lewis and Wilding, 1984; Millar, 1979; Colquhoun, 1981; Smith, 1987b, 1992; Chiles, Alluisi and Adams, 1968; Blake, 1971;). Day-of-week effects have also been documented. For instance Testu and Clarisse (1999) found superior performance on Thursday as opposed to Monday, however consideration of these factors in detail is beyond the scope of this thesis. Clearly, Smith (1992) was correct when he concluded that the whole testing situation must be considered when interpreting time-of-day effects in performance.

1.9. Contemporary Theories of Diurnal Performance Fluctuations

Subsequent to the unidimensional theory of arousal being deemed as an inadequate full and complete explanation of time-of-day effects in performance, alternative explanations have been proposed. For instance, Hockey and Colquhoun (1972) suggested that tasks with a large memory load might be expected to exhibit an inverse relationship to temperature, while immediate processing tasks may follow a more direct relationship. However, rather than interpret this 'memory effect' in terms of arousal, Hockey and Colquhoun (1972) suggested that this effect is directly linked with changes in basic processes like metabolic rate. It is suggested that a higher metabolic rate increases the speed of decay in short-term memory, but increases the rate of processing in other types of task. Anderson and Revelle (1994) suggested that impulsivity may be a good predictor of rate of change in arousal level, this was found not to be the case and alternatively it was proposed that susceptibility to lapses in attention are mediated by phase differences, related to the impulsivity of the individual, in the diurnal arousal rhythm. Subsequently, assuming arousal and alertness are synonymous Owens, Macdonald, Tucker, Sytnik, Minors, Waterhouse, Totterdell and Folkard (1998) investigated whether alertness could predict time-of-day effects in performance. It was found that although alertness was quite a good predictor of simple perceptual-motor speed measures, it was less useful in predicting some other measures. The time-of-day trend for all measures was different to that for alertness. It was concluded that alertness can successfully predict changes in some measures of performance, but extrapolating to other performance measures should be done with caution. Finally, Coyle (1989) investigated a "levels of control" approach that postulated that task performance is governed at an executive level and at a lower level where automised decisions are taken. Results indicated that control at the upper level becomes worse over the day while control at the lower level improves. It was concluded that this approach provides a superior framework for the analysis of the cognitive processes involved in human performance.

Other theories that have received more consideration include that of Monk and Leng (1982) who found evidence to suggest that two underlying arousal rhythms may be responsible for time-of-day effects in performance rather than just one. It was suggested that there was one arousal rhythm that was parallel to the temperature rhythm which was responsible for performance on low memory load tasks and a second arousal rhythm that peaked three hours earlier, which was responsible for performance on medium and high memory load tasks. This was argued to be consistent with Folkard and Monks' (1981) multioscillator model of circadian performance fluctuations which was proposed to account for differences in rates of adjustment of rhythms following shiftwork for example. Folkard, Wever and Wildgruber (1983) stated that such a model assumes "any given circadian rhythm to be jointly controlled by two endogenous oscillators. The first is thought to be relatively immune to exogenous factors and to control the temperature rhythm, while the second is thought to be more influenced by exogenous factors and to have the major role in governing the sleep/wake cycle" (p.1). Interestingly, Folkard et al (1983) found evidence that working memory performance was controlled "by a previously unidentified oscillator with an autonomous period of about 21h" (p.1). Similarly, Monk, Weitzman, Fookson, Moline, Kronauer and Gander (1983) proposed that different relationships sometimes seen between the time-of-day trend seen for temperature and that seen for performance may be a result of control of performance by either a 'strong oscillator', which presumably controls core body temperature, or a 'weak oscillator', which presumably controls the sleep-wake cycle. Further, Monk, Weitzman and Fookson (1983) also believed that performance is mediated by an endogenous circadian process which is related to the temperature rhythm and also by a homeostatic process that is coupled with the sleep/wake cycle. Johnson, Duffy and Dijk (1992) found support for this notion by using a forced desynchrony protocol (where the temperature rhythm and the sleep/wake cycle are forced to desynchronise) to attempt to separate the effects of the circadian timing system and time since awakening. Johnson et al (1992) also found that when the sleep/wake cycle was suspended in this way and data collection proceeded into the night, a parallelism between short-term memory performance and temperature emerged.

Folkard et al (1983) and Folkard, Marks, Minors and Waterhouse (1985) thus proposed that there exists one or more 'performance oscillators'. Supporting the multioscillator model, Folkard and Monk (1984) concluded that circadian rhythms in performance were probably "not all mediated by a single underlying factor such as arousal" (p.190), further they stated "it would appear that at least two underlying factors (one of which may correspond to arousal) are responsible for mediating such rhythms and that these factors may be controlled by different oscillators" (p.190).

In sum, as stated by Carrier and Monk (2000), contemporary theories of subjective alertness and performance efficiency believe these to be controlled both by a homeostatic system (amount of hours of wakefulness) and by a circadian timing system (for example, Borbely, 1982; Monk et al, 1983; Dijk, Duffy and Czeisler, 1992; Folkard and Arkerstedt, 1992; Johnson et al, 1992). Such models postulate that diurnal fluctuations in performance efficiency are believed to be a result of an interaction between these two systems, where performance on a task may deteriorate over the day because of the number of hours spent awake (homeostatic process) or because the circadian timing system results in a sub-optimal state at which to perform the task, or because of a combination of these. Alternatively, performance efficiency may remain stable over the day because the circadian timing system counterbalances the result of having been awake for a number of hours (Carrier and Monk, 2000).

As Carrier and Monk (2000) concluded "performance rhythms do not appear to be the simple *direct* result of circadian changes in either mood or physiology. The understanding of the mechanisms underlying different diurnal fluctuations during waking hours (without suspending the sleep/wake cycle) will require dissection of the individual effects of homeostatic and circadian influences on performance efficiency. This will not be a simple task since current research suggests that these processes vary with task parameters (for example, cognitive load) and individual characteristics (age, chronotype, level of practice)" (p. 727). Therefore, "there is still much work to do before one can understand which performance tasks will show different time-of-day effects and what the mechanisms are that underlie these differences" (p. 721).

1.10. Summary

There is now much evidence supporting the notion that performance efficiency and other psychological factors, such as mood, are subject to circadian variations. Much work has been carried out in an attempt to determine what processes underlie this circadian variation. Originally, Kleitman (1939; 1963) believed time-of-day effects in performance to be due to an increase in thought processes resulting from increases in body temperature over the day, thus Kleitman (1939; 1963) believed the relationship between body temperature and performance to be a causal one. However, this notion of causality has subsequently been rejected (for example, Colquhoun, 1971; Smith, 1992). Colquhoun (1971) proposed that arousal level was the underlying factor mediating performance rhythms. It was proposed that basal arousal was a single underlying process that mediated the relationship between temperature and performance, suggesting that performance levels could be predicted from temperature readings. The arousal model can account for many different performance trends according to the level of complexity (or memory load) involved in the task. Folkard (1983) believed the arousal theory of time-of-day effects in performance to be 'remarkably useful' (p.267) and stated that it has been successful in predicting interactions between the time-of-day and other factors such as stressors that are believed to influence arousal, and in general it has been consistent with the results of various time-of-day studies. Yet there are problems with this theory. This becomes apparent when consideration is given to the dissociation of temperature and performance for particular personality types. Consequently, Folkard (1983) concluded that this unidimensional model of arousal is 'probably over-simplistic' (p.268). Subsequent theories have suggested that a homeostatic process and input from a circadian timing system may underlie circadian variation (for example, Borbely, 1982; Monk et al, 1983; Dijk et al 1992; Folkard and Akerstedt, 1992; Johnson et al, 1992).

Further, the relationship seen between diurnal variation in performance and body temperature that gave rise to the arousal theory is now known to be true only for certain tasks and individuals completing those tasks. Task demands, especially memory load, exert a large influence over the trend in performance seen across the day. Even in simple, non-memory loaded tasks the parallelism between body temperature and task performance is dependent on the individual completing the task (Folkard, 1983). Accordingly, Folkard (1983) concluded that the best time-of-day to perform a task depends on the nature of the task itself. As Smith (1992) stated, it may in some circumstances be advantageous to schedule some tasks for completion in the morning and others for completion in the afternoon. Smith (1992) concluded that is important to consider performance, not only in relation to the physiological state of the individual, but in the context of the whole testing environment. It is important to examine as many factors as possible relevant to performance as opposed to considering time-of-day effects alone. However, as Folkard (1983) concluded, "it is undoubtedly the case that such (time-of-day) effects exist, that they are relatively unavoidable, and that they have important practical, as well as theoretical, implications for the study of human performance efficiency" (p.268).

1.11. Time-of-Day and Icon Use

Here, what little work has been done to date on the effect of the time-of-day on a persons' ability to complete a task using symbology will be considered, before moving on to consider the aims of this thesis. Consideration of the general topic of icons will be the concern of chapter 2.

1.11.1 Time-of-Day and Icon Interpretation

To date, only one study has examined the effects of the time-of-day on the completion of an icon search task. McFadden and Tepas (1997) examined the effects of the time-of-day and task demand on simulated highway sign recognition. Reaction times and accuracy rates varied significantly according to the time-of-day and experimental condition (that is, the number of pieces of information to process, or memory load). A significant interaction was also found between these main effects, where reaction times were slowest at 0900 in the high task demand (high memory load) condition and were slowest at 1500 in the low task demand condition (low memory load) condition for both participants. Fastest responses occurred at 1200 in the high task demand condition for both participants and also at 1200 in the low task demand condition for participant one, while fastest response times in the low task demand condition occurred at 0900 for the second participant. As they stand the results are not consistent with previous memory load literature. However, McFadden and Tepas (1997) believed that it may have been possible that participants limited their search to a specific part of the icons in condition A (high task demand/memory load) and that this therefore meant that condition A actually represented a lower memory load than condition B. Mean response times and accuracy rates support this. Indeed, if the results are reversed then their findings are consistent with literature finding that working memory performance peaks at 1200 (Laird, 1925; Owens et al, 2000; Folkard, 1975) and that in the case of the data of participant number two, the peak in working memory performance occurs earlier on more highly loaded memory tasks (Folkard et al, 1976; Folkard, 1983).

The aim of this thesis is to expand on the exploratory study conducted by McFadden and Tepas (1997). The extent to which an individuals' ability to complete an icon search task is influenced by the time-of-day at which performance is assessed will be examined. Results will be considered in terms of the influence of the time-of-day on different types of memory and memory load. If the time-of-day influences the ease with which icons and visual displays can be interpreted and understood, then the results will have wide applicability to everyday tasks such as driving, to the design of cockpits and to optimising human performance.

Chapter 2
Icon Use and Performance

2.1. Introduction

The wide use of computers on a daily basis is now becoming a fact of life that was envisaged some years ago (Mills, 1967). However, for the information technology society to reach this stage much painstaking research has been undertaken that has enabled the expert and novice alike to interact with computers with a degree of ease. Much of this research has centred on the disciplines of human computer interaction, cognitive psychology and cognitive ergonomics. Within these disciplines solutions have been found for the many problems that have faced interface users. To understand human performance in computing environments, it is essential that we have an understanding of the variables that influence performance and devise ways of measuring these variables (Miller, 2001).

2.2. Mental Models

Much cognitive psychology research has concentrated on 'mental models' of the interface user. 'Mental models' refers to the process of acquiring new knowledge and reorganising existing knowledge. Craik (1943) proposed that the notion of mental models means that when an individual interacts with external events these events are translated into internal, mental models that are then manipulated to interpret the system in front of them and then provide an action. Essentially then, users are not interacting with the computer as such, but with their mental models of them (Manktelow and Jones, 1987). However, there are individual differences in the development of mental models, with the mental models of experienced users differing from those of novice users. We shall return to this later.

Icons were first employed on an interface in the Xerox 'Star' office application (Smith, Irby, Kimball and Verplank, 1982), the interface design was based on the metaphor of an actual office (for example, file management systems were represented by an icon of a filing cabinet). The value of this type of interface is that the icons would then correspond with user's pre-existing mental models. Rogers (1989) stated that icons work as pictorial representations of "aspects of an interface metaphor" (p.105). The most well known of these metaphors is the desktop metaphor, originating with the Xerox application, where the interface has been designed to represent a desktop. Objects that would normally be found on an office desk such as folders and files are used in icons to represent computer files and file storage (Rogers, 1989).

Icons are increasingly used to depict a wide range of functions from the signs we see everyday on public toilets to complex representations of processing in chemical plants.

So we have the notion then, that usable interfaces should contain metaphors that can interact with users pre-existing mental models. Smilowitz (1996) highlighted the fact that user interface guidelines frequently encourage the use of metaphors in icon design, suggesting that software packages should build upon the real world knowledge of the user by using concrete metaphors to make systems more usable. In spite of this, little research has supported the concept that metaphors facilitate user performance. Thus, Smilowitz (1996) asked participants to perform a series of tasks using both metaphoric and non-metaphoric interfaces. In Experiment 1 a library metaphor generated better performance than a nonmetaphoric screen. In contrast, Experiment 2 found a travel metaphor did not yield a performance advantage over a nonmetaphoric interface. Further experiments revealed metaphors were most effective when they formed the essential component of the concept being represented as opposed to being composed of different metaphor elements. Finally, it was found that maximising the degree of similarity between the metaphor and the concept it represented contributed to the effectiveness of the metaphor. Such findings may explain why the library metaphor was advantageous over a non-metaphoric display while the travel metaphor was not.

Further evidence that the use of metaphors may not provide the whole key to usability comes from a study by McDougall, Curry and de Bruijn (2001). They examined the notion that the context of a graphical interface, or the visual information that a person is presented with, can partly determine the nature of the mental models that the user develops. It was found that the mental models of the user do depend on the nature of the graphical/visual information on the interface but these mental models do not depend as much on adopting visual metaphors as was previously thought but are more dependent on semantic/articulatory distance (the difference between an icon and what it means (Blankenberger and Hahn, 1991); see section 2.3.3).

2.3. Icons

So we have the notion that metaphors on a usable interface must be presented in iconic (or pictorial) form, yet Smilowitz (1996) found some evidence that although a metaphor generally resulted in better performance than a nonmetaphoric interface, icons did not necessarily contribute to the advantage held by the metaphoric interface.

This raises questions about the extent to which icons really are advantageous in contributing to usable interfaces. This is complicated by different views about just what an icon is. Wood and Wood (1987) believed an icon to be any image that can be used to represent a concept or object. Horton (1994) stated "other terms for simple visual images are symbol, sign and signal" (p.3) and that "the term icon is often used as a synonym for any small visual symbol" (p.2).

2.3.1. Advantages of Icons in Comparison to Text

Conventional command line interfaces, for example MS-DOS, are difficult to assimilate and hence are difficult to remember, additionally they are not useful in helping the individual visualise the underlying operations (Barker, Najah and Manji, 1987; Lodding, 1983). It is expected that an expert who had taken time to understand and learn the interface would best understand interfaces such as these. Icon-based interfaces are therefore thought to have considerable advantages over command-line interfaces.

Icons represent abstract concepts better than text (Rogers, 1986; Rogers and Oborne, 1987) and are more efficient in their use of human cognitive characteristics such as visual recognition (Muter and Mayson, 1986; Hemenway, 1982), memorability (Banks and Flora, 1977; Rogers, 1989) and interpretation (Blankenberger and Hahn, 1991). Icons have the potential to be universally applicable, fully able to overcome the limitations encountered by verbal languages (Rogers, 1989; Stotts, 1998). Arend, Muthig and Wandmacher (1987) concluded that computer functions depicted as text commands are searched and selected for very slowly. Arend et al (1987) stated that icons are pictorial or graphical signs that are related to the concepts they represent in terms of a similarity between the picture constituting the icon and the concept or object that the picture is intended to represent. As a result, it is probable that icons possess a smaller articulatory distance to their respective meanings than text commands. Arend et al (1987) stated that this would explain why icons can generally be identified faster than text commands and why they are less likely to cause error, but this may only be true in the case of novice users: Text commands, *if* they can be understood are more precise than icons, but icons are more likely to be understandable and therefore represent less of an articulatory distance than text commands for novices. It is interesting however, to note that users have expressed a preference for icons over text even when they did not facilitate usability (Kacmar and Carey, 1991).

Wiedenbeck (1999) has examined the role of icons directly. He compared the learning of a program where the interface used buttons with text labels, icons or a combination of icons and text labels. During the first session, performance was best using the label-only and icon-label interfaces. Performance on the icon-only interface was found to be poorer during the first session with an improvement during the second session. Retention of skill between the first session and the delayed session was worse for the icon-only interface however this effect did not persist long. Subjects perceived the icon-label interface as the easiest to use, while their perceptions of usefulness was greater for the icon-only and icon-label interfaces than for the label-only interface during the first session. Thus, it would appear that the interface must employ icons, a label alone is not enough, however it seems that the combination of an icon and a label is probably best, the more information that is available the better. Heck (1996) examined whether pictorial or textual representations were more effective at communicating their intended meaning in situations that varied in their visual complexity. Results showed that icon representations were more effective than textual ones in communicating their meaning in a variety of situations. It was found that icons were continually selected more quickly than the textual representations regardless of the complexity of the interface in which they were displayed. This supports the conclusion that icons are a vital part of a usable interface.

A number of reasons have been proposed for the popularity of icons. Stotts (1998) suggested that an iconic interface is more familiar and simplifies the system. Additionally, as good icons reduce dependence on text, they are invaluable for those with reading disabilities such as dyslexia, for those whose first language is not English and for those who are illiterate (Horton, 1994).

Horton (1994) stated that icons yield their meaning quickly, represent visual and spatial concepts, save space and speed up the search process. He continued by saying that as icons are readily and easily learned, "once we have learned the unique shape of an icon, we remember it reliably and recognise it immediately" (p.5) and that as humans "we remember visually encoded concepts better than those encoded verbally" (p.5). Additionally, Horton (1994) stated that icons are remembered better than text, firstly because icons are more visually distinct from each other than words, secondly because when we see a symbol we name it and remember the name and the symbol, thus icons are stored in memory visually and verbally but words are only stored verbally, and finally because the visual images are stored in several forms and are tightly linked to one another and other forms. He explains that good icons are essential when individuals must act quickly. Hence icons are frequently used when fast and accurate responses are required, for example on road traffic signs. However, it must be noted that Horton (1994) does not cite research evidence to support his assertions, even though such evidence can be found. For instance, studies of road traffic signs have found that iconic signs can be read at twice the distance and in half the time as text signs. Kline, Ghali and Kline (1990) for example tested young, middle aged and elderly people and for all three age groups they found that icon signs could be seen at much greater distances than text signs and this was found to be more pronounced at dusk. Interestingly, Kline et al (1990) found no age differences in the ability to understand the signs but they did find that there was immense variability between the icons in the degree to which they could be understood. This highlights the importance of good icon design.

The concept of mental models can help us consider the question why icons appear to be better than text. There are advantages of using stimuli that makes use of an individuals' pre-existing world knowledge, for instance by allowing an individual to adapt their pre-existing mental models rather than making them develop new ones the complexity of the system is reduced. This is exactly what modern operating systems allow the user to do by using illustrations of objects that are often found in an office such as files, folders, in/out trays and waste paper bins (Streitz, Liesser and Wolters, 1989; Rohr and Keppel, 1984). To elaborate, Rohr and Keppel (1984) proposed that the reason why iconic interfaces are better than verbal command sets is because they can be constructed in a way that gives the user a better chance of implicitly acquiring a model of the system structure than verbal command sets would allow. This proposal is based on the fact that there are specific areas of information presentation in existence, where complicated information can be presented in more condensed and wholistic manner by using icons.

Thus research has shown that icons have advantages over text. Unfortunately, this has led to the blanket use of icons in some applications where they are not always appropriate, to the extent that their use can actually have a detrimental effect on system usability (Rogers, 1988). It is the design or characteristics of the icons, and not just the icons per se, that are important in order to allow greater usability. Consequently, we need to know when icons should be used and what form they should be presented in (Rogers, 1989). So although there is clear evidence that icons usually enhance performance, more research is needed into what extent performance is influenced by icons and into what characteristics of icons improve performance (Stotts, 1998).

McDougall et al (1996) found that the usability of icons depended on various factors such as complexity, user experience and task demands. They noted that other factors such as stress and fatigue may also affect user performance. Therefore, it appears that we must consider both the nature of the icons and the nature of the individuals using them (for example, their level of experience). The effects of a wide range of icon characteristics on performance have been considered. These include icon complexity, concreteness, meaningfulness, familiarity and distinctiveness (McDougall et al, 2001) as well as icon uniqueness, ambiguity and dominance (Goonetilleke, Shih, On and Fritsch, 2001). The effects of icon characteristics on user performance that are pertinent to the research that follows will now be considered. Subsequently, individual differences in user experience will be considered.

2.3.2. Icon Complexity

McDougall et al (1996) defined complexity as "the amount of detail or intricacy in the symbol" (p.9). McDougall and de Bruijn (1999b) stated that it is important to distinguish between visual complexity, conceptual complexity and display complexity. It is stated, "visual display complexity refers to the visual complexity of the display as a whole and, in particular, to the manner in which it is organised" (p. 14). For example it is noted that while cockpit displays may often use simple icons, the cockpit as whole is visually complex. It is also stated "conceptual complexity refers to the complexity of the meanings activated by displays. It may be completely independent of visual complexity" (p.14).

There are many guidelines in existence stressing the importance of using simplicity in icon design, Easterby (1970) stated, "...symbols should be as simple as possible. Fine detail makes no contribution to unambiguous and rapid interpretation of a symbol" (p.157). Byrne (1993) found that simple icons were easier to discriminate than complex ones when performing a search task, suggesting that icon complexity is the key. McDougall et al (1996) also suggest that simple symbols improve user performance because they can be identified more easily as they can be discriminated using relatively few features. As a result, search times are reduced when trying to locate simple symbols in a display. Complexity effects have been found to influence search efficacy even when users are experienced (McDougall et al, 1996). They suggested that this was because search is a basic visual process. The use of complex icons will therefore lead to slower response times even when users are experienced.

One of the potential problems interface designers face when trying to reduce complexity on displays, is that this appears to conflict with the need for icons to be pictorial or concrete (in order to produce visual metaphors). Often researchers have increased the complexity of icons in order to make them more pictorial (for example, Stammers, George and Carey, 1989; Arend et al, 1987). Indeed, Stammers (1990) pointed out that although previous research has suggested that pictorial or, concrete, icons are easier to use than those which are not, there are limitations on how pictorial icons can be when this means that they are also more complex. He continues, stating that some evidence suggests that complex icons may not give rise to the best performance and that simpler icons may give better results. However, it is apparent from icons currently in use and from previous research (Curry, McDougall and de Bruijn, 1998; McDougall et al, 2001) that icons do not need to be complex to be concrete and that concreteness and complexity can be varied orthogonally (see Figure 1).

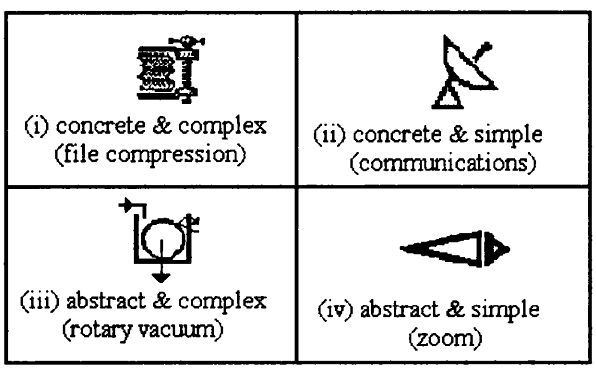


Figure 1: Example of the orthogonal nature of concreteness and complexity, taken from Curry et al (1998).

The potential confounding of icon concreteness and complexity was perhaps most apparent in a paper written by Garcia, Badre and Stasko (1994). They devised a metric to measure icon concreteness, in their words "a quantitative measure for abstractness based on the complexity of the icon" (p.191). The metric was strongly related to perceptions of complexity and involves adding up the number of features present in the icon. For example, features might include the number of horizontal, vertical and diagonal lines, arrowheads, arcs, lines, letters and special characters. Superficially at least, the metric appeared to provide a good match to subjective measures of icon concreteness. However, Garcia et al (1994) validated their metric using icons that had been employed in previous research (Stammers et al, 1989; Rohr and Keppel, 1984; Rogers, 1986; Arend et al, 1987) in which the researchers had simply added more detail to icons to produce a concrete or pictorial set. Figure 2 shows an example of the problem. Rogers (1986) produced concrete and abstract icons, but concrete icons were much more detailed so that, when features were added up, concreteness scores were higher than abstractness scores. This however, was primarily the result of their complexity not their concreteness.

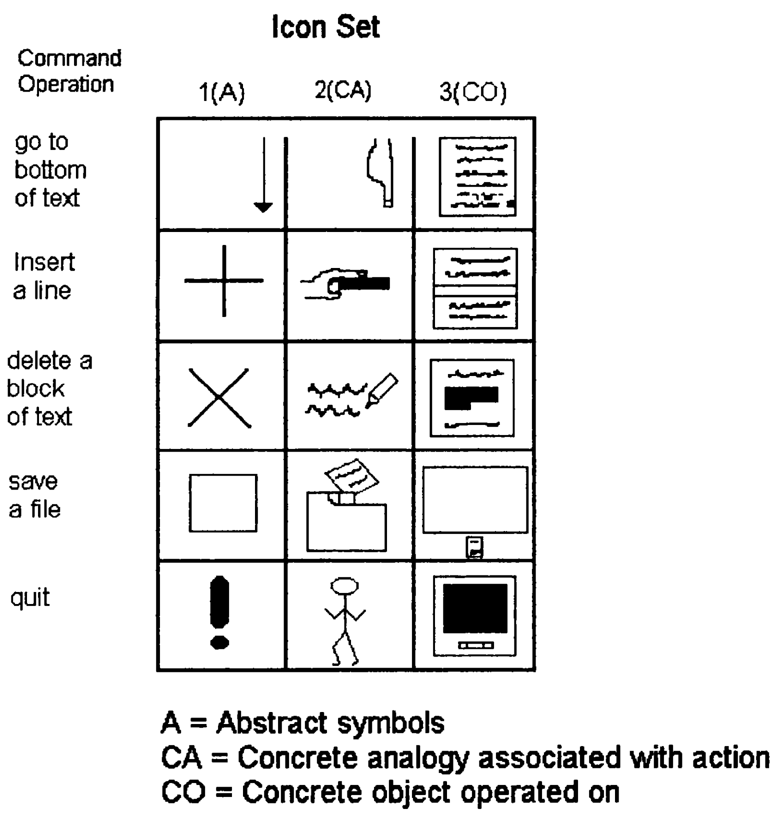


Figure 2: Examples of icons that have been increased in complexity to increase concreteness, taken from Rogers (1986).

McDougall, Curry and de Bruijn (1999a) examined the possibility that complexity and concreteness had been confounded in previous studies. They obtained subjective ratings of icon concreteness and icon complexity. McDougall et al (1999a) believed that a strong correlation between the two ratings would support the idea that complexity and concreteness are parallel characteristics. On the other hand, if no correlation was found the suggestion would be that the two dimensions have indeed been confounded in previous research. There was little relationship between concreteness and complexity ratings suggesting that they are two separate dimensions that were confounded in the previous studies used by Garcia et al (1994) to validate their data. This means that *both* simple and complex icons can be concrete or abstract (see Figure 1). Interestingly, McDougall et al (1999a) did find a strong correlation between the metric produced by Garcia et al (1994) and subjective measures of visual complexity (although *not* with subjective measures of concreteness). This provides further support for the idea that Garcia et al's (1994) measure is one of complexity, not concreteness.

In a review of the literature, McDougall et al (1996) stated that little work had been done on the effect *icon* complexity may have on user performance, while much effort had concentrated on the effect *display* complexity may have. They highlighted the existing controversy between the use of simple icons compared to the use of more complex icons. For example Byrne (1993) found that the most important factor in search processes is the type of icons used. In an experiment where simple, complex and blank (described as 'non-visual' p.448) icons were used, simple icons outperformed complex and blank icons, further if the target icon in the simple icon condition was unique to the display, the advantage of this condition over the others was considerable, while if half of the display consisted of similar targets this effect disappeared. It was concluded that if icons are to be advantageous then they must be "simple and easily discriminable" (p.452) complex icons are not likely to enhance performance because they are "difficult to discriminate quickly" (p.452). Hence it seems that Bryne (1993) connected simplicity with discriminability and therefore wished to imply that icon simplicity and icon distinctiveness are interrelated.

Although Garcia et al (1994) suggested that complicated icons have advantages, this was only because they were more concrete. As discussed earlier, Garcia et al's (1994) metric involved counting up the icon features, the problem with this is that concreteness and complexity were confounded (see Figure 2), this meant Garcia et al (1994) were effectively suggesting that concrete icons were, in part, effective because of their visual complexity and the extra detail they contained. What seems more likely is that the pictorialness of the icons meant they generated better performance in Garcia et al's (1994) task, which required associations with meaning. On the other hand Byrne (1993) used a visual search task that did not involve meaning, clearly then simplicity was advantageous to the demands of Byrne's (1993) task where users had to search for a target; basic visual search is profoundly influenced by the complexity of the icon (McDougall et al, 2001). Thus, the conflicting conclusions regarding complexity given by Garcia et al (1994) and Byrne (1993) can be explained in terms of task requirements, the extra complexity shown in Garcia et al's (1994) study was advantageous due to the task in hand, likewise the simplicity used in Byrne's (1993) study was advantageous due to the characteristics of his task (McDougall et al, 1996).

Indeed, findings from McDougall et al (2001) suggest that task demands are important in determining the role of complexity. In their first experiment, users were required to search within an array for a given icon. In this search task, search times for complex icons were slower than for simple icons. The influence of icon complexity on response times did not diminish with increased user experience (Interestingly, concreteness had no effect on response times in this simple search task). In a second experiment, the researchers introduced a 'matching' element to the task. Participants were given a function label and asked to match it to one of the icons in the array. Thus, they not only had to search the array but they also had to derive the meanings (or functions) of the icons in the array to obtain a successful match with the function label given. In this task, icon concreteness also influenced response times in addition to complexity. They concluded that this was because concrete icons could be matched more quickly with the function label as meaning was more easily derived from concrete (pictorial) rather than abstract icons. It was interesting to note however, that the effects of icon concreteness were short-lived and that once the icon set was learned, and users knew the meaning of the icons, no differences were found between concrete and abstract icons. Yet the difference in response times between complex and simple icons remained. Taken together these findings suggest that icon complexity has long lasting effects closely linked to the basic visual processing of an array while icon concreteness has relatively short-lived effects linked to deriving meaning from the icons on an interface.

In sum, it can be concluded that the effects of icon complexity concern the time taken to search a display and respond appropriately. Icon complexity is associated with search efficacy; increasing visual complexity increases visual search time. Simple icons should be used if users are completing a visual search of a display where prompt responses are required. As the effects of visual complexity are associated with basic visual processing, it is not affected by experience and applies to novice and experienced users. The use of simple icons is less important if the task involved does not include a strong search element, if the task is not time critical or if the task mainly involves accurately understanding the interface/determining a meaning or function (McDougall, Curry and de Bruijn, 2000).

2.3.3. Icon Concreteness

Curry et al (1998) defined concreteness as "the degree of pictorial resemblance that an icon bears to its' real-world counterpart" (p.1590). Rogers (1989) classified icon types into four groups: *resemblance icons* that "depict the underlying referent through an analogous image," (p.110; for example, the road sign warning of falling rocks, see Figure 3a); *exemplar icons* which were described as "a typical example for a general class of objects" (p.110; for instance, the knife and fork sign used in information signs to represent restaurant services, see Figure 3b); *symbolic icons* which were described as conveying "an underlying referent that is at a higher level of abstraction than the image itself," (p.110; for instance, the image of a glass with a crack in it represents the 'fragile' concept, see Figure 3c); and *arbitrary icons* which are described as bearing "no relationship to the referent and hence the association must be learned" (p.110; for instance, the icon for a biohazard, see Figure 3d).

Rogers (1989) stated that the effectiveness of an icon in conveying its' intended meaning is dependent upon the degree of mapping (articulatory distance) between the physical object and the function the icon represents. This relates to the concept of concreteness. A concrete icon is closely mapped to the object it is intended to represent in that it is usually more pictorial and is often metaphoric so that it can correspond to the users' mental models (Rogers, 1989; Smilowitz, 1996; Steitz et al, 1989; Rohr and Keppel, 1984; Stammers et al, 1989; Arend et al, 1987; Garcia et al, 1994; McDougall et al, 2001). Alternatively, an abstract or arbitrary icon can be used where "there is no connection between the physical representation and the underlying

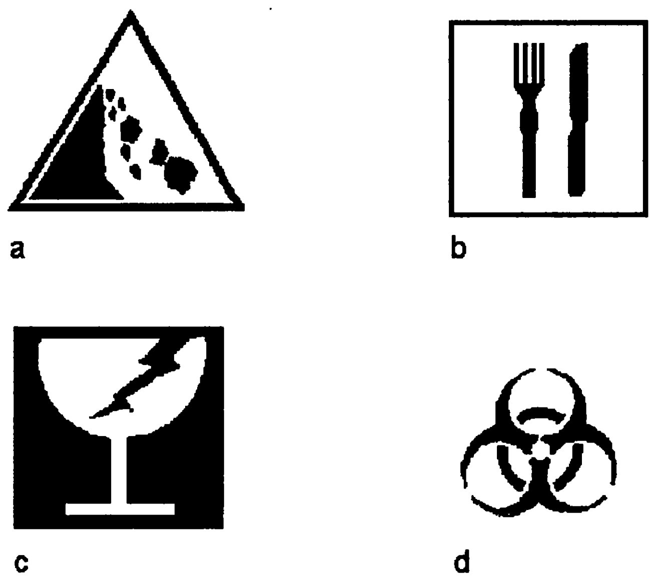


Figure 3: Examples of four groups of icons, classified by and taken from Rogers (1989).

representation" (Rogers, 1989 p.111). To elaborate, icon concreteness refers to the degree to which an icon actually looks like the object or function that it represents. If an icon does not resemble the object or function that it represents then it is referred to as an abstract icon. As noted earlier, McDougall et al (2000) found evidence to suggest that concreteness effects mainly influence the initial understanding of the icons' meaning. Hutchins, Hollan and Norman (1986) also suggested that icon sets that are comprised of concrete icons, minimise articulatory distance and therefore can be learned more readily than icons that are more abstract. However, the relationship between pictorialness and articulatory distance is not necessarily as simple as this. Consider two icons that are used to represent computer functions, a picture of a printer and a picture of a tortoise. The printer represents the print function and therefore the icon directly represents the function, minimising articulatory distance, on the other hand, the tortoise infers a slow function, thus increasing articulatory distance. Consequently, equally concrete icons can differ in articulatory distance

(McDougall and de Bruijn, 1999b). Furthermore, Rogers (1989) stated that there is a good case for abstract icons in that they "are an effective form of coding because they have no prior associations. Hence, once a person has learned to associate the meaning between the physical form and the underlying referent of a symbol they should have no subsequent difficulty understanding the meaning of it" (p.111). However, because the fit between the picture and its function in abstract icons is not very good, the connection between the icon and its' referent is not easy to learn or remember.

Evidence to date suggests that concrete icons are more readily learned and recognised than abstract icons. Jones (1983) found that participants who were creating pictograms of abstract verbal concepts demonstrated a preference for using concrete as opposed to abstract representations, even when the concreteness values were very low. Garcia et al (1994) found that concrete icons were identified more easily than abstract icons and Rogers (1986) found that the extent to which icons were initially meaningful was dependent "on the directness of the mapping between the pictorial representation and referent" (p.600). Stammers et al (1989) asked subjects to match an icon to a function shown in a display, their response times and accuracy of response were recorded over five trials. An advantage for concrete items appeared at first but disappeared over the subsequent trials. Similar findings are reported by McDougall et al (2000) and Green et al (1990) who suggest that the difference in performance between abstract and concrete icons dissipates with experience. Finally, Stotts (1998) examined current word processing icons and found that 'graphically concrete' icons (icons that display a realistic representation of the object of reference) and 'functionally representative' icons (icons that "have an inherent and direct relationship between the object of reference and its intended action" (p. 454)) were recognised more quickly and more accurately than 'graphically abstract' icons (icons that display only an outline of the object of reference) and 'functionally arbitrary' icons. As a result, Stotts (1998) proposed that icons should be created to look like the object of reference as much as possible.

However, specific task demands may also play an important part in usability. For example, assuming articulatory distance and concreteness are synonymous, Blankenberger and Hahn (1991) examined the influence of articulatory distance on performance in a 'search and select' task. They found that articulatory distance influenced reaction times where the icon screen positions were randomised but not when the icon screen positions were fixed, a finding supported by Green and Barnard

(1990). Furthermore, Stammers (1990) found icon appropriateness to be an important factor in icon usability. He asked the same subjects in two separate groups to identify redesigned icons in terms of abstractness-concreteness or to rank them on an appropriateness scale. It was found the appropriateness ratings predicted identification performance. It was also found that the redesigned icons were considered to be more concrete and more appropriate. He concluded that user appropriateness assessments could predict subsequent usability, but in contrast to Stotts (1998), stated that it is an oversimplification to suggest a guideline to make icons as concrete as possible.

In sum, it can be concluded that a concrete icon closely resembles the object or function that it is intended to represent. The closer this relationship the more effective the icon is likely to be. Concrete icons generate better performance than abstract icons initially, however, this advantage is short-lived and once the meaning of the abstract icons has been learned, the difference between concrete and abstract icons dissipates. Concreteness and complexity have been confounded in previous research where the complexity of icons has been increased in an effort to make them more concrete. Subsequent research has shown that these dimensions are not related and both simple and complex icons can be concrete or abstract (see Figure 1).

2.3.4. Other Icon Characteristics

Although icon concreteness and icon complexity are regarded as important icon characteristics, a number of other characteristics have been considered including icon distinctiveness.

As indicated above, it appears that the distinctiveness of the icons designed plays a significant role in their usability. Research has shown that distinctive icons are more easily and more quickly recognised (Aspillaga, 1996) and are easier to find quickly in displays in which they may be presented with other icons (Fisher and Tanner, 1992; Byrne, 1993), furthermore they are not as easily confused with other icons (Magyar, 1990). McDougall, de Bruijn and Curry's (2000) explored the features which make icons distinctive, participants in their study identified features that they considered made the icons more distinctive. These were as follows: (1) icon simplicity or complexity (2) meaningfulness (3) darkness versus lightness (4) the presence of global features (e.g. symmetrical, existence of emerging pattern) or local features (specific features of icons) (5) size. Arend et al (1987) found that icon distinctiveness can be associated with concreteness. They found participants responded to abstract icons more quickly than concrete icons when performing a search task, suggesting that abstract icons are more distinctive. Further, response times showed the abstract icons could be searched in parallel, while the concrete icons had to be searched sequentially. However, efforts had been made to maximise the distinctiveness of the abstract icons by employing global features (that is, shape, colour, size), while the concrete icons employed local features (that is, lines and structures within figures) to ensure representativeness and small articulatory distance. The concrete icons used seemed to be less distinctive and more complex than the abstract icons. Nonetheless, this demonstrates that icon distinctiveness can override the advantage normally seen for concrete icons in a search task.

Distinctiveness can be enhanced through the use of visual and semantic contrasts (McDougall et al, 2000), in support of this Nasanen, Ojanpaa and Kojo (2001) found that when contrast increases, search time decreases. However distinctiveness contrasts have been found to be complex (McDougall et al, 2000). McDougall et al (2000) obtained subjective ratings of icon distinctiveness under different conditions. They found simple icons presented against an array of complex icons to be very distinctive and against an array of simple icons concrete-complex but not abstract-complex icons were found to be distinctive. Further, distinctiveness ratings of targets varied with changes in the concreteness of the background array, it was found that distinctiveness ratings of targets were higher when the array they were set against was abstract. To expand, distinctiveness ratings for concrete icons when the array consisted of abstract icons were high, but this was not so when an abstract icon was shown against a concrete array. Thus, effective contrasts were only achieved when concrete icons were shown against an array consisting of abstract icons. McDougall et al (2000) concluded that two types of contrast can be created: (1) a visual contrast – mainly concerns differences in icon complexity (2) a semantic contrast – mainly concerns differences in icon concreteness. These contrast effects are not symmetrical and are only effective when simple icons are presented in a complex array, when concrete-complex icons are in a simple array and when concrete icons are in an abstract array.

This research therefore suggests that icon distinctiveness can enhance user performance and can override normal complexity and concreteness effects. However, distinctiveness effects are complex and visual/semantic contrast effects are not symmetrical. Distinctiveness is a feature that may prove important in the research that follows although its' effects were not specifically tested. The effects of icon gestaltness however were specifically tested and this will now be considered.

Any stimulus that results in the perception of wholeness can be considered to be a gestalt. Sekular and Blake (1994) stated that the Gestalt principles highlight the factors that are thought to produce a perception of wholeness, briefly these are: proximity, which is the tendency of stimuli to group together to form one perceptual unit; similarity, which is the tendency for items that are similar, for instance in terms of lightness/darkness, orientation and size, to be grouped together; closure, which is the tendency to perceive contours that are close together as one; and good continuation, which is related to closure and is where neighbouring stimuli are grouped together on occasions where they may be potentially connected.

Theories of visual attention centre on a limit in our ability to see several things at once and are related to the gestalt principles. These theories are: discrimination-based theories, which suggest that there is "a limit on the number of separate discriminations that can be made" (Duncan, 1984 p.502); space-based theories that suggest that there is "a limit on the spatial area from which information can be picked up" (Duncan, 1984 p.502); and object-based theories, that suggest that there is "a limit on the number of separate objects that can be perceived simultaneously" (Duncan, 1984 p.501). Duncan (1984) stated that the evidence to support the discrimination and space based theories is not strong, for this reason and because it is likely that it will be the object-based theories that will be most relevant in the experiments that follow, attention here will focus on object-based theories.

Neisser (1967) proposed two stages exist in the object-based theory of visual perceptual analysis: in the preattentive stage the field is divided into separate objects according to the gestalt organisational rules of good continuation and proximity for instance. The second stage is that of focal attention which analyses an object in greater detail. It is believed that the first stage is parallel across objects presented simultaneously, while the second is serial and it is this stage that imposes the limit on how many objects we can see at once.

Treisman, Kahneman and Burkell (1983) found support for the object-based theory, finding that when participants had to identify a word and the location of a break in the outline of a box simultaneously, the task was completed more efficiently when the box surrounded the word rather than being presented at the opposite side of the display. Merikle (1980) found further support for the object-based theory, finding that reporting letters presented in one row or column was superior when the letters were grouped by colour. These studies, concerned with visual grouping, appear to give at least some support to the object-based theory and are clearly related to gestalt principles. Indeed, Duncan (1984) proposed that "gestalt grouping processes serve to package information preattentively into chunks, which then serve as units for focal attention" (p.502). In support of this, Duncan (1984) conducted an experiment using two overlapping objects, a box with a line through it. It was found that two judgements about the same object can be done simultaneously without compromising accuracy, but two judgements about different objects cannot. Duncan (1984) believed that this supports the theory that parallel preattentive processes divide the field into separate objects and focal attention then follows that deals only with one object at a time. Finally, Goldsmith (1998) examined whether location (space-based) or perceptual object (object-based) processes were used for feature integration in visual search. It was found that the search process was most efficient when the features of the search were connected to the same perceptual object (object-based) than when they were connected to different perceptual objects at the same location (space-based), lending firm support to object-based theories. However, the object-based advantage was found to depend on the discriminability and density of the stimulus, grouping strength and also hierarchial object structure.

Boersema and Zwaga (1996) stated that a common problem in everyday life is finding direction signs in visually noisy environments where many signs and adverts are displayed, for example at an airport. Feature integration theory (Treisman and Gelade 1980) suggests that this problem can be solved by giving the direction sign a unique feature that is not present elsewhere in the same environment. Similarly, the attentional engagement theory of Duncan and Humphreys (1989, 1992) suggests that finding a direction sign can be made easier by reducing the similarity between the sign and objects found in the same environment. So these theories, suggest a 'pop-out' effect of the target stimuli due to its' uniqueness or dissimilarity to the rest of the environment (Boersema and Zwaga, 1996). It is important to note that these theories appear to be related to the concept of distinctiveness. Treisman (1986) assumed that parallel processing is used during the early stages, where visual processes scan the entire field simultaneously, while serial processing is used later on where the individual details are given attention. Next, elements that were detected very early on during processing are indentified, these elements can be detected without close scrutiny anywhere in the field; these elements 'pop-out' from their surrounding environment. The features that were found to 'pop-out' were for example, colour, curves, tilted lines, target contrast, proximity, length and number. Finally, it is important to note a study by Boersema and Zwaga (1996) that found that 'pop-out' of easy targets from their surrounding environment did not occur during the first few trials in a task, but found practice was required for this 'pop-out' effect to occur.

In this thesis the extent to which icons could be seen as coherent objects rather than as sums of parts was explored. Given Goldsmith's (1998) theoretical perspective one might expect icons that can be seen as a coherent whole to be responded to more quickly. Furthermore, icons which are coherent wholes might be more distinctive and easier to discriminate between, creating a 'pop-out' effect to some extent.

2.3.5. Measuring Icon Characteristics

The metric created by Garcia et al (1994) has already been considered. It was apparent that although Garcia et al (1994) intended to measure icon concreteness they were in fact were measuring visual complexity. This problem illustrates some of the difficulties associated with measuring icon characteristics appropriately. McDougall and de Bruijn (1999b) made one of the first attempts at addressing this issue directly. Participants in their study were asked to provide subjective ratings of a number of icon characteristics including concreteness, visual complexity, meaningfulness, familiarity and semantic distance (McDougall et al's, term for articulatory distance). Their methodology was almost identical to that employed by Paivio and others to measure the concreteness, familiarity and meaningfulness for example, of words (for example, Paivio, Yuille and Madigan, 1968; Gilhooly and Logie, 1980; Snodgrass and Vanderwart, 1980). They found that ratings for icon concreteness, meaningfulness, familiarity and semantic distance were all inter-correlated. This can be understood by considering that when an icon is concrete it closely depicts a real world object, which would make it meaningful and familiar and minimise semantic distance. Interestingly, the visual complexity ratings obtained for the corpus of icons did *not* correlate with concreteness, meaningfulness, familiarity or semantic distance, but *did* correlate with the scores of each icon on Garcia et al's (1994) metric. This not only adds weight to the notion that Garcia et al (1994) were really measuring visual complexity, but also suggests that this characteristic is fundamentally different in some way. Indeed, research to date suggests that icon complexity is an index of the basic visual search process involved in a task while the other icon characteristics are primarily associated with access to meaning (McDougall et al, 2000).

McDougall et al (2000) pointed out that creating a metric to measure distinctiveness is difficult, especially if the distinctiveness of an icon changes according to the context in which it is presented or if an icon's distinctiveness rests on its' semantic features. Despite this there have been previous attempts to measure distinctiveness using perceptual discriminability (Geiselman, Landee and Christen, 1982). Although perhaps measuring icon characteristics need not be so difficult, Sears, Jacko, Brewer and Robelo (1998) found evidence to suggest that simply the ability of an interface user to identify the functionality of an icon from the graphic image alone may be an effective way of evaluating the quality of new icon designs.

In conclusion, it certainly appears that the solution to measuring icon characteristics and generating usable icons lies in the complexity and concreteness features, as all other features appear to be related to concreteness. However, icon distinctiveness has the potential to become a very prominent feature, if not in individual icon design, then certainly in interface design, however this is a complex effect the measurement of which does not appear to be any simpler.

2.4. User Experience

Generally, as a person becomes experienced in a task their performance will improve beyond that seen for a novice. Much of the research into this difference between experts and novices has focused upon areas such as chess (Chase and Simon, 1973; de Groot, 1965), mathematics (Lewis 1981), physics (Larkin 1983; Chi, Glaser and Rees, 1983) and computer programming (Bateson, Alexander and Murphy, 1980; Adelson, 1981c). From such research the notion has arisen that the mental models of experts are more advanced (they have a greater understanding of the problem) and more importantly, are abstract. Meanwhile, the mental models of novices are less mature and more concrete (DiSessa, 1983). Lamberti and Newsome (1989) examined differences between expert and novices in problem solving tasks, using two types of questions, some that required abstract organisation of information and some that required concrete organisation of information. The 'high skill' participants were found to be significantly faster in responding to the abstract questions whilst the 'low skill' participants were significantly faster on the concrete questions. So mental models change with experience and therefore it is not surprising to learn that there are differences between experts and novices in their interaction with computers (Stotts, 1998). Such research illustrates the need for the interface design to be compatible with the cognitive capacity of the user. Here, how human performance on computer related concepts in general is influenced by user experience will be considered before specifically looking at the influence of experience on concreteness and complexity.

Moyes and Jordan (1993) pointed out that although much work has illustrated how the most successful icons closely illustrate their actual function, little work has considered how this effect may change over time and with different levels of experience. They proposed that an individual's performance when using a new interface would represent a normal learning curve where performance improves with practice (see Figure 4). Three components were considered central to usability – learnability, experienced user performance (EUP) and guessability. It is stated that these components explain how performance on a new interface is likely to improve before reaching an asymptotic level (that is, when a maximum level of performance is reached) forming a learning curve. The three components were defined as follows: guessability – "the measure of the cost to the user involved in using an interface to perform a new task for the first time. The lower the cost the higher the guessability (cost can be measured either in terms of time, errors or effort)" (p.51); learnability – "the measure of the cost to the user in reaching some reasonable level of performance on a task" (p.51); experienced user performance – "the measure of the cost to the user of performing a task when they have reached a relatively steady level of performance. Again, the lower the cost, the higher the EUP." (p.52).

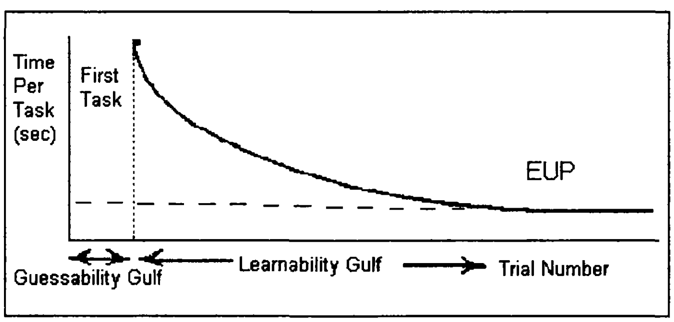


Figure 4: The 'typical learning curve' including the guessability, learnability and experienced user performance components, taken from Moyes et al (1993).

Moyes and Jordan (1993) postulated that when an individual is using a new interface the degree to which the icons can be guessed is important, but over time as the individual becomes more familiar with the interface the guessability (concreteness) of the icons becomes much less important. Indeed, Stammers et al (1989) and McDougall et al (2001) have found support for this, finding performance advantages for concrete items initially, with this wearing off as the number of trials increases.

Similarly, Stotts (1998) found experienced users to be faster and more accurate overall than novice users. When graphical abstraction was concrete both experienced and novice users responded more correctly and faster. This is in support of Garcia et al (1994) and Blankenberger and Hahn (1991). Moreover, with graphically concrete icons experienced users were nearly twice as fast as novices and responded more correctly. It was also found that experienced users showed a higher rate of increase in reaction time between concrete and abstract icons compared to novice users. Evidence to suggest that as long as icons are functionally representative and graphically concrete they will be recognised quicker was found throughout the conditions in Stotts (1998) study. Regarding accuracy, it seems functional representation is the mediator, as when icons are functionally representative they are recognised accurately more often than when icons are functionally arbitrary.

Further, Blankenberger and Hahn (1991) found that when icons with different articulatory distances were randomly positioned on a screen, articulatory distance had an effect on reaction time, but when icons were selected from fixed screen positions this effect of articulatory distance was not seen. However, the authors found evidence to suggest that articulatory distance had little effect on the performance of experienced users.

In contrast to concreteness and articulatory distance, the effects of icon complexity, in tasks with a search component, do not reduce with experience. McDougall et al (2000) found that the effects of visual complexity persisted even when users became experienced because analysing visually complex material requires basic visual search processing and this remains constant irrespective of experience.

Thus differences in performance between experienced and new users are well documented in general areas of psychology as well as in icon usability. The primary points to note here with regards to experience in icon usability are as follows: performance differences between concrete and abstract icons are initially profound, but as the user gains experience and learns the meaning of the icons this difference gradually reduces; for complexity effects however, where performance is best using simple icons, are persistent, it seems likely that because this process involves basic visual search, experience has no effect.

2.5. Visual/Iconic Memory

One aspect that appears to have been neglected in experimental investigations into the ability of an individual to successfully complete a task using symbology is visual/iconic memory. This is potentially very important, as effects concerned with complexity and concreteness in a task that does not involve visual/iconic memory may be very different than in tasks that do necessitate the images to be remembered.

Iconic memory was investigated by Coltheart (1980) who believed that a person may have knowledge about the visual characteristics of an image only when an unrecoded representation of the image was present, this was supported by the fact that information about an image is available to a person beyond the physical cessation of the image. Coltheart (1980) thought two forms of iconic memory existed, one that was more durable than the other, he believed that in order for information about an image to be reported, it had to be transferred to the more durable form of iconic memory. Evidence was reported for a limited capacity of iconic memory of four or five items.

However, it is often the case that not only does the icon image needs to be remembered, but it's meaning must be memorised too. Rogers (1989) stated that a reason why iconic interfaces are easier to learn and remember may be that users are not required to recall command labels, rather they are only required to recognise and select the appropriate icon from an array of others. This minimises cognitive load on memory. Rogers (1989) concluded that this means that the advantages arising from using icons rather than other interface forms must be due to "differences between the content of the displayed information" (p.114). Rogers' (1988) study found icon mapping of the most direct form (that is, the most concrete) to be most effective, this was reflected by fewer requests for help and the most correctly identified icons in the memory task. Similar results were found for a label task, but more errors in the memory test were apparent. Also, memory for the meaning of the icons improved over time, while it remained the same for the label task. Therefore, subjects could perform tasks equally well for both icon and label tasks, but more difficulty in remembering what the labels meant was apparent, while subjects had little difficulty in remembering what the icons meant. Rogers (1989) concluded that there is a difference in the way the information is used and stored for the two forms of communicating information. Subjects could easily recall information when icons were learned but not when labels were learned; this can be explained in terms of the type of information coded. Rogers (1989) referred to Paivio's (1971) dual coding theory that suggests that the meaning of an icon is likely to be better remembered because the pictorial information is likely to be stored in visual and verbal memory stores, presumably this was thought to improve memory because the material could then be accessed in two different ways rather than just one. Indeed, Zhou (2000) suggested that the mapping from the mental space in memory storing information regarding computer function to the mental space in memory storing the icon image, is the key to the recognition of computer functions as represented by the icon image on the interface. A series of experiments found the strength of the connection between the two mental spaces in memory were not symmetrical; participants could recognise currently-used icons more effectively than they could generate representations of new icons. Moreover, it was found that recognition strength accounted for the most variation in performance.

In sum, as discussed earlier, the meaning of icons may be remembered better than the meaning of text commands. Rogers (1989) explained this in terms of the information being stored in both visual and verbal memory stores. Yet the effect of visual memory in icon tasks has been largely ignored, perhaps because human-interface interaction does not usually require the user to remember the icons (Rogers, 1989). However, it appears that iconic memory has a limited capacity (Coltheart, 1980) suggesting that the visual/iconic memory component involved in a task may be important. Consequently, visual/iconic memory will be considered as an aspect of task demands in the experiments that follow.

2.6. Summary

It was thought that visual metaphors were the key to icon usability so that the icons would then interact with a persons pre-existing mental models. Indeed, icons have been shown to be superior, in terms of ease and speed of performance, to text (Arend et al, 1987; McDougall et al, 1996) and to be preferred by users (Kacmar and Carey, 1991). This superiority has been attributed to an icons' ability to interact with pre-existing mental models that serves to reduce the complexity of the system (Streitz et al, 1989; Rohr and Keppel, 1984).

In spite of this, the visual metaphor is not as important as was first thought but the characteristics of the icons have been found to be important in determining usability. Two primary icon characteristics were considered here: complexity and concreteness. Much research has highlighted the importance of simplicity in icon design (Byrne, 1993; McDougall and de Bruijn, 1999b). Yet the complexity of icons has often been increased to make them more concrete, or pictorial, (Stammers et al, 1989; Arend et al, 1987) to ensure they represent effective visual metaphors. Consequently, a problem is encountered where pictorial/concrete icons enhance user performance while complex icons impair it. Such confounding of these characteristics was very apparent in Garcia et al's (1994) research. However, symbol complexity is not related to concreteness (McDougall et al, 1999a), as a result *both* simple and complex icons can be concrete or abstract, thus realistic icons with no increase in complexity are possible (Curry et al, 1998).

It appears that the complexity effect where response times are consistently slower when using complex stimuli remains even in experienced users. This is likely to be due to the search process involving basic visual processing that cannot be improved with practice. This is in contrast to concreteness effects, here response times are initially faster for concrete icons and are slower for abstract icons, but this effect disappears as users become more experienced and in doing so have learned the meaning of the icons (McDougall et al, 2001). It seems simplicity is important where response times are vital, while simplicity becomes less important when the primary task is to understand the icon or where search is not involved in the task (McDougall et al, 1999b; McDougall et al, 2000).

Another icon characteristic that affects usability is distinctiveness. Research has suggested that distinctive icons improve user performance (Aspillaga, 1996). Even large articulatory distances and can be overridden by increased icon distinctiveness (Arend et al, 1987). However, distinctiveness effects have been found to be complex, but can be enhanced through visual and semantic contrasts, for example showing concrete icons against an array of abstract icons. However, contrasts are not symmetrical; showing abstract icons against an array of concrete icons does not work (McDougall et al, 2000).

The gestaltness of an icon, or the degree to which it produces a perception of wholeness, also appears to be of importance. In the experiments that follow it is believed that the object-based theories of visual attention will be most useful. The object-based theory has actually received a lot of empirical support (for example, Triesman et al, 1983; Goldsmith, 1998).

Finally, an aspect of the task that seems to have been largely neglected in literature to date is the effect of visual/iconic memory on tasks employing symbology, albeit, human-interface interaction does not usually require the user to remember the icons (Rogers, 1989). Rogers (1989) found that users remembered the meaning of icons better than the meaning of text commands, this was explained in terms of the information being stored in visual and verbal memory stores. Moreover, it seems the capacity of iconic memory is limited (Coltheart, 1980). Consequently, it appears that the visual/iconic memory component of a task using symbology may be an important feature of the task demands.

2.7. Time-of-Day and Icon Use

An aspect of icon interpretation and the time taken to respond to them that has received little attention is the effect of the time-of-day. Only one previous study has attempted to examine this (McFadden and Tepas, 1997, see Chapter 1). Differences in performance resulting from differences in the time-of-day have been found in many cognitive functions, such as memory (see Chapter 1), it is clear that if similar differences exist in the interpretation of, and time of response to, icons then the implications for time and safety critical applications, such as air traffic control, would be immense. As discussed in Chapter 1, it is exactly this void that the research for this thesis will attempt to fill.

Chapter 3

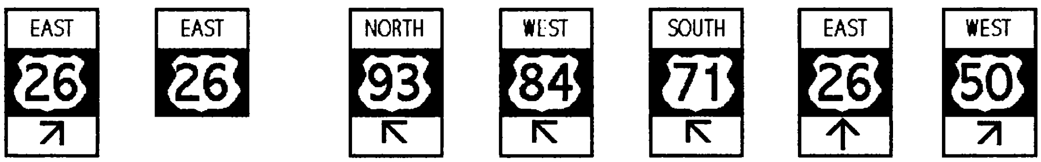
Experiment 1: A Replication of McFadden and Tepas 1997

3.1. Introduction

As noted previously, the use of symbology in everyday life has become increasingly common over recent years. Research has shown that there is a great deal of variability in the degree to which icons can be understood (for example, Rogers, 1986, 1989; McDougall et al, 2000; Stotts, 1998). Moreover, research has shown that a number of factors can affect the ease and speed with which icons are used. These include the characteristics of the icons themselves such as their complexity, concreteness and distinctiveness, and also individual differences in experience, between those who use icons (Byrne, 1993; Rogers, 1989; Aspillaga, 1996; Moyes and Jordan, 1993). For instance, it is widely recognised that simple icons result in faster responses than complex icons and this has been attributed to simple icons being identified more easily as they can be discriminated using relatively few features (Byrne, 1993). Further, this complexity effect has been found to be unaffected by changes in user experience (McDougall et al, 2001).

One factor that has been found to affect human performance is circadian variation. The effects of circadian fluctuation in areas such as mood (for example, Monk et al, 1985) and memory (for example, Monk et al, 1978; Baddeley et al, 1970; Folkard, 1975) are well-documented, yet the pervasive effect of this biological phenomenon has been largely neglected in consideration of icon interpretation. It seems possible that circadian variation may cause a slowing of response times and/or an increase in error rates in the completion of an icon task at certain times of the day. Such changes have important implications for optimising human performance and improving safety in many occupational areas such as air traffic control.

Only one study has examined this possibility to date. McFadden and Tepas (1997) demonstrated the existence of a time-of-day effect in the use of road sign symbology that varied according to task demands (that is, the number of pieces of information to process, or memory load). Two male subjects were asked to perform a road sign recognition task four times a day (at 0900, 1200, 1500 and 1800) for eighteen (non-consecutive) days. Their subjects searched through computer screen displays of American road signs for a specified target sign. The subjects indicated whether or not the target was present. Each display consisted of 5 signs (see Figure 5c). The target sign was presented for 15 seconds before the display from which the subjects had to choose appeared. The authors used a high and low task demand condition, which used complex and simple icons respectively. In the low task demand condition (see Figure 5b) the target sign showed a United States interstate route number and a direction (that is, north, south, east or west). In the high task demand condition (see Figure 5a) an arrow (left, right or straight) was also shown.



(a) Complex (b) Simple (c) Example of Icon Array

Figure 5: Icons Used by McFadden and Tepas (1997)

The subjects completed both conditions during each session. Participants completed 200 trials in each condition and 20 practice trials making a total of 420 trials per session. Error rates, reaction times and body temperature were recorded.

McFadden and Tepas (1997) found that response times and error rates varied significantly according to the task demands and the time-of-day, and there was a significant interaction between the two. The slowest response time occurred at 0900 in the high task demand/memory load condition (condition A) and at 1500 in the low task demand/memory load condition (condition B). While fastest response times occurred at 1200 in the high task demand/memory load condition for both participants and also at 1200 in the low task demand/memory load condition for participant 1, while the fastest response times in the low task demand/memory load condition occurred at 0900 for the second participant. However, they noted that according to the memory load literature (for example, Folkard 1983), these results should be reversed. McFadden and Tepas (1997) believed that condition A (originally, high task demand/memory load) allowed participants to search for the arrow only (allowing participants to look at one, rather than three, icon features). This meant that condition A constituted a lower task demand (or memory load) than condition B (originally, low task demand). This makes McFadden and Tepas' (1997) study consistent with findings in the memory load literature. McFadden and Tepas' (1997) findings of an increase in response times at 1500 is also consistent with previous literature that has reported a post-lunch dip in performance (for example, Lenne et al, 1997). However, much research has demonstrated the afternoon dip in performance to be flexible (for example, Craig et al, 1981; Smith and Miles, 1986b, 1987b, Smith et al, 1990).

Memory and the Arousal Model

Memory literature suggests that the early morning is associated with superior immediate memory and this is because of the low levels of arousal associated with this time-of-day (Berger, 2000; Winch, 1912; Monk et al, 1978). On the other hand, research examining working memory, has found performance on these tasks to peak at 1200 (Laird, 1925; Owens et al, 2000; Folkard, 1975). This time of optimal performance falls between the decreasing time-of-day trend for immediate/short-term memory and the increasing time-of-day trend seen for simple immediate processing tasks such as visual search (Folkard, 1983). Performance on working memory tasks has also been found to peak earlier in the day on more highly loaded memory tasks (Folkard, 1976).

Memory load is, in fact, an important factor in determining the exact diurnal performance trends observed. Indeed, the arousal model, which has been a common explanation for time-of-day performance trends, asserted that an arousal rhythm mediated temperature and performance trends and that performance could be predicted from temperature. The arousal theory suggested that a high level of arousal is advantageous for low memory load tasks, while a low level of arousal is advantageous for high memory load tasks (Colquhoun, 1971). Accordingly, assuming arousal increases over the day, it has been reported that a low memory load task shows an evening peak in performance (Blake, 1967b; Folkard and Hill, 2002), while a high memory load task shows a morning peak (Folkard and Monk, 1980). This all suggests that low arousal is a prerequisite for optimal performance of a complex (high memory load) task.

Monk (1982) developed Colquhoun's (1971) theory and proposed the existence of two arousal rhythms that mediated performance, one which paralleled the temperature trend and was responsible for performance on low memory load tasks and another that peaked three hours earlier and was responsible for performance on medium and high memory load tasks. Other more contemporary theories suggested that a homeostatic process (time since awakening) and input from the circadian timing system both determine diurnal performance trends (for example, Monk et al, 1983; Monk et al, 1989; Dijk et al, 1992; Johnson et al, 1992; Folkard and Akerstedt, 1992).

However, as Carrier and Monk (2000) concluded, separating the individual effects of both these factors is not easy, especially when diurnal trends are sensitive to manipulations of task demands.

Indeed, all this shows that exact task demands are very important in determining time-of-day performance trends, accordingly, previous work that has found time-of-day effects to be influenced, for example, by changes in motivation (Smith, 1992; Chiles et al, 1968; Blake, 1971). Consequently, Smith (1992) concluded that the whole testing situation must be considered in the interpretation of diurnal performance trends.

Body Temperature

McFadden et al (1997) also found body temperature to vary according to the time-of-day, with temperature rising over the course of the day. Indeed this has been long known and it has been suggested that the temperature rhythm somehow mediates or even causes performance rhythms (for example, Kleitman, 1939, 1963). Indeed, Johnson et al (1992) found that if the sleep/wake cycle is suspended and data is collected into the night then a parallelism between short-term memory and temperature emerges. However, more recent work has failed to find any relationship between temperature and performance (see Owens et al, 2000). Indeed, Campbell (1992) noted that experiments that have found a parallel between performance and temperature often used boring vigilance tasks that did not require higher level processing. When higher level processing is involved, such as memory load, the relationship between temperature and performance is not so straightforward.

In consideration of other possible physiological indices of performance, Carrier and Monk (2000) stated that although the temperature rhythm and the rhythms of cortisol and melatonin secretion have the same controlling pacemaker as performance rhythms and have been found to correlate, these correlations were not very high. Given the low correlations observed, Carrier and Monk (2000) concluded that performance rhythms are probably best considered as being controlled relatively independently by the circadian timing system resulting in a trend that sometimes coincides with other trends seen for other variables "without necessarily being directly mediated by any particular physiological rhythm" (p. 726). This is borne out by the fact that the performance decrements reported during the post-lunch period (for example, McFadden and Tepas, 1997; Kleitman, 1939, 1963; Owens et al, 2000;

Lenne et al, 1997) are not mirrored in the temperature readings taken at that time-of-day (Colquhoun 1971).

In the experiments that follow tympanic membrane (eardrum) temperatures were taken to confirm that participants demonstrated normal rhythms and to examine any relationship between temperature and performance. This method was used for a number of reasons. Firstly, the eardrum shares its' blood supply with the hypothalamus (the 'temperature control centre' of the brain) thus tympanic membrane temperature should give a better indication of core body temperature, as oral temperature readings can be influenced greatly by hot and cold drinks for example, while rectal temperatures lag behind changes in core body temperature. Tympanic membrane temperatures remain relatively unaffected by such factors (www.mypharmacy.co.uk). Also perhaps tympanic membrane temperatures can be considered to be a better reflection of brain temperature and it is brain temperature that is perhaps most important if temperature does indeed relate to performance on cognitive tasks. Finally, tympanic membrane thermometry is now widely used by medical practitioners and as such is likely to be highly reliable. Furthermore, tympanic membrane temperature is very easy to measure.

Experiment 1

Experiment 1 was designed as a partial replication of McFadden and Tepas' (1997) study. There are however, some important differences between the two studies that are designed to address some possible problems in McFadden and Tepas' (1997) study. In McFadden and Tepas' (1997) study, only two participants were used. This study, in contrast, used a total of fifty participants. Secondly, McFadden and Tepas (1997) presented the target sign for a period of 15 seconds. This is rather a long period of time, particularly since a driver would not have the opportunity to look at a road traffic sign for this period while in a moving vehicle, therefore this study presented the target for a much shorter period of 2 seconds. Additionally, an extra testing time was added in Experiment 1, McFadden and Tepas' (1997) study tested until 1800, but this study tested until 2100. This was to examine any time-of-day trends that may persist into late evening. Each participant in this study completed their test sessions in one day, while McFadden and Tepas' (1997) participants were tested over 18 days. Finally, a new set of icons was designed that paralleled McFadden and Tepas' (1997) icon set in terms of the number of pieces of information presented. McFadden and Tepas' (1997) icons were not used because they were American road signs and there were no British equivalents (see section 3.2.1. for an example of icons used).

3.1.1 Aims

Experiment 1 examined the following:

1) Whether ability to complete an icon task was influenced by the time-of-day, both in terms of accuracy of response and response time.
2) Presuming a time-of-day effect was found, whether simple and complex icons showed different diurnal performance trends.
3) If the type of icon, simple or complex, influenced reaction time overall regardless of the time-of-day.
4) Whether temperature values showed any relationship to the diurnal performance rhythms in these tasks.
5) Whether differences in experience with icons (manipulated between experimental conditions) influenced performance.

On the basis of research to date, it was expected that response times and accuracy rates would vary as a function of the time-of-day and according to memory load. It is important to note at this stage that throughout all experiments there were three aspects of the tasks that each contributed to the memory load involved. These were:

(a) *The visual memory component.* Whether the target icon disappeared from screen during the search and therefore had to be remembered (increasing memory load), or whether the target icon remained on screen during the search and therefore did not need to be remembered (decreasing memory load);

(b) *The difficulty of response.* Tasks that required a 'yes'/ 'no' response increased memory load as participants not only had to decide if an icon was present or not but then also had to recode their response into a key press. Tasks requiring a mouse click response decreased memory load as the icon was always present and participants simply had to click on it with the mouse;

(c) *The difficulty of icon discrimination.* Some icons were more difficult to discriminate than others either due to their complexity (with simple icons decreasing memory load), degree of wholeness (with gestalt icons decreasing memory load) or their distinctiveness (with distinctive icons decreasing memory load). If a visual memory component was also involved then the difficulty of icon discrimination was increased further.

In this way task difficulty and/or complexity have been equated with memory load. As all experiments, with the exception of Experiments 3 & 4, required participants to hold information in memory while the task was carried out, these tasks were thought to involve working memory. It was expected therefore, that the timing of peak performance would vary in much the same way as other working memory tasks used in previous research, that is peak performance would occur at 1200 with the peak occurring earlier on more highly loaded memory tasks.

The memory load involved in Experiment 1 was high, the task involved: (a) a visual memory component; (b) a difficult 'yes'/ 'no' response; (c) non-distinctive icons that were difficult to discriminate. However, the complex icons used in Experiment 1 were likely to create a higher memory load than the simple icons, as they consisted of three rather than two components that each had to be remembered. Consequently, in this high task demand/memory load condition, the fastest response time was expected to occur early in the day at 0900, while in the low task demand/memory load condition (simple icons) the fastest response time was expected to occur later in the day at 1200. A post-lunch dip was expected to appear in the tasks at 1500. Further, simple icons were expected to result in more accurate and faster responses. Whether temperature values would show a relationship to any diurnal performance rhythm found remained to be seen, it was difficult to predict the outcome here with much controversy existing in the literature. Finally, it was expected that faster performance would result from increased experience, which was manipulated between experimental conditions.

3.2. Method

Experiment 1 used a visual search paradigm where a target icon had to be identified as either being present in, or absent from, an array of 5 icons.

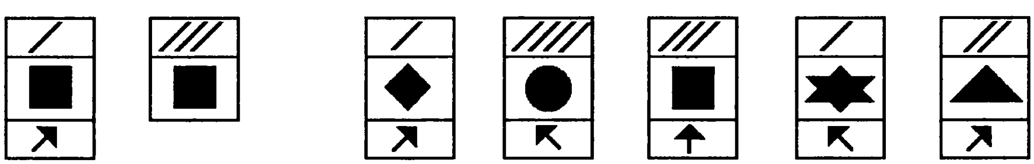
3.2.1. Participants

Fifty undergraduates and postgraduates from the University of Wales Swansea were participants. Nine were male and forty-one were female. Five males and twenty females completed the unequal experience condition while four males and twenty-one females completed the equal experience condition. The age range of participants in the unequal experience condition was 18 – 25 years; the mean age was 19 years 9 months (standard deviation, 2 years 3 months). The age range of participants in the equal experience condition was 18 – 24 years; the mean age was 19 years 10 months (standard deviation, 1 year 9 months). Some participants obtained course credit of 3 hours for taking part while others received a payment of £10.

3.2.2. Materials & Apparatus

Icons were presented using a self-paced computer program that moved on to the next trial as soon as the participant had responded and automatically recorded reaction times and accuracy rates. The visual search task was presented on Pentium 166 MHz computers. The screen settings on all computers were set to 1024x768 pixels. Reaction times were measured using the systems' multimedia timer, allowing measurement to within 1 millisecond resolution. A Braun Thermoscan thermometer (Model IRT 3520) was used to measure participants' temperatures before and after each test session.

All icons were black and white. The information conveyed by the icons was designed to parallel the information seen in the icons used by McFadden and Tepas (1997), effectively replicating their experiment. Each icon had three possible pieces of information and each piece had the same number of possible levels as those icons used by McFadden and Tepas (1997). Thus the complex icons (see Figure 6, a) included three pieces of information: distance (400, 300, 200, 100 yards); route (square, triangle, circle, diamond, star); and direction (straight on, left, right). The simple icons (see Figure 6b) included just two pieces of information; distance and route. Each piece of information included in these icons was taken from existing UK road traffic signs, additional route (star) and distance (400 yards) information was also used to parallel McFadden and Tepas' (1997) study.



(a) Complex (b) Simple (c) Example of Icon Array

Figure 6: Icons Designed for Experiment 1

3.2.3. Design

Participants were divided into ten groups of five, one group of five participants were tested at a time. Each participant was tested once at each of the following times: 0900; 1200; 1500; 1800; 2100. Each participant also completed a practice session at one of the test session times. The practice session a participant attended, and the order of testing thereafter, was counterbalanced using a Cyclic Latin Square (see Table 1) to ensure the administration of the experimental conditions was balanced across participants. Each participant's sessions were completed consecutively within 24 hours.

Group	*Practice*	*Test 1*	*Test 2*	*Test 3*	*Test 4*	*Test 5*
1	2100	0900	1200	1500	1800	2100
2	0900	1200	1500	1800	2100	0900
3	1200	1500	1800	2100	0900	1200
4	1500	1800	2100	0900	1200	1500
5	1800	2100	0900	1200	1500	1800

Table 1: Cyclic Latin Square

3.2.4 Procedure

In each trial a target icon was shown for a period of 2 seconds before disappearing from screen. Immediately following this, the array of 5 distractors appeared on screen (see Figure 7). Participants were instructed to respond as quickly and as accurately as they could at all times.

The participant had to decide whether the target icon was present in the display or not. If the target icon was present, participants were instructed to press the 'Q' key, if the target was not present, participants were instructed to press the 'P' key. The 'Q' key was labelled 'Y' for 'yes' and the 'P' key was labelled 'N' for 'no' to avoid confusion. Participants were asked to place their fingers on these keys for the duration of the experiment. Timing began when the icon array appeared and finished once participants had responded by pressing a key. The next trial began and the whole procedure repeated. The program did not respond to erroneous key presses. No feedback was given regarding correctness of response. If no response was given within 6 seconds, the program moved on to the next target automatically.

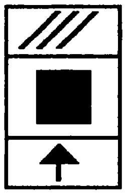


Figure 7(a): Screen 1/Complex icons, *target icon appears for two seconds.*

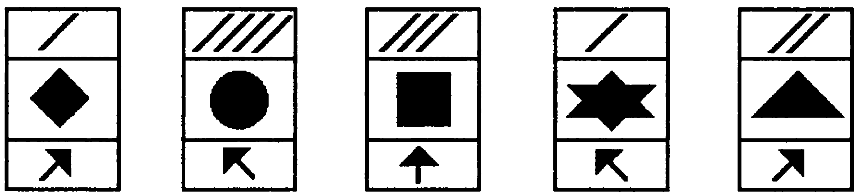


Figure 7(b): Screen 2/Complex icons, *target disappears & distractor array appears for six seconds, in this case target icon is present among the distractors. Participant should press the 'Q' key labeled with 'Y' for 'yes'.*

Immediately after the first set of trials was completed, the Experimenter began the second set of trials and the same procedure ensued. Eardrum temperature readings were taken before and after each test session and were taken from the right ear on every occasion. The procedure was identical at every test session.

Conditions, blocks of trials & randomization of icons

Two conditions were used, these differed in the amount of experience given to participants completing each condition (unequal experience versus equal experience). Half of the participants had unequal experience of both icon types, while the other half had equal experience of both icon types. Participants were allocated to either the unequal or equal experience condition on an alternating basis. The condition where participants were given an unequal amount of experience with complex and simple icons constituted a strict replication of McFadden et al's (1997) study. In the other condition, experience of both complex and simple icons was equated. Both groups of participants completed 180 trials using a set of 60 complex icons. However, the participants in the unequal experience condition also completed 180 trials using a set of 20 simple icons resulting in them having three times more experience with the simple icons than the complex icons. In order to equate the amount of experience participants had with each icon set, the participants in the equal experience condition only completed 60 trials using the same set of 20 icons. In both conditions an array of 5 icons was used consisting of the target icon and 4 distractor icons in the present trials and of 5 distractor icons in the absent trials. Selection of the targets and the distractor arrays was randomised so that each had an equal chance of being sampled and the participants saw a different set of stimuli during each test session.

<u>Unequal Experience</u>

Complex Icons

There were 180 trials using a set of 60 icons. 60 icons were designed to enable each combination of icon features to be presented. The target icon was present among the distractor array in 120 trials and was absent in 60 trials (that is, the present to absent ratio was 2:1). Each icon acted as a target 3 times in 1 set of 180 trials, however each icon was only visible among the icon array twice because there were 60 absent trials. The distractor icons were chosen at random and the location of these distractors (and targets) in the array was randomised. Each icon was presented 13 times as a distractor. Thus the search task was designed to ensure that all icons were presented the same number of times as both targets and distractors in order to ensure that opportunities for gaining experience with icons was equated across the icon set.

Simple Icons

There were 180 trials using a set of 20 icons. 20 icons were designed to enable each combination of icon features to be presented. The target icon was present among the array in 120 trials and was absent in 60 trials (ratio 2:1). Again, the search task was designed to ensure that participants' experience with individual icons in the set was equated. Each icon acted as a target 9 times in a set of 180 trials, however each icon was only visible among the distractor array six times because there were 60 absent trials. Each icon was presented 39 times as a distractor in 180 trials. The location of these distractors (and targets) in the array was randomized.

The unequal experience condition paralleled McFadden and Tepas' (1997) study, where the participants had different levels of experience with complex and simple icons. Each icon type was presented across a set of 180 trials, thus participants completed a total of 360 trials at each test session.

Equal Experience

Complex Icons

The way in which complex icons were presented in this condition was identical to the unequal experience condition. Each participant was given 180 trials, which meant that each icon acted as a target 3 times and each icon was presented 13 times as a distractor.

Simple Icons

Although participants were given 180 trials for both the simple and complex icons in the unequal experience condition, there were fewer icons in the simple set which meant that they were shown more frequently in a 180 trial cycle. Here, participants were presented with *60* trials using the same set of 20 simple icons as used previously. The target icon was present among the array in 40 trials and was absent in 20 trials (that is, a 2:1 present:absent ratio). Each icon was presented 3 times as a target in a set of 60 trials, however each icon was only visible among the distractor array twice as there were 20 absent trials. Each icon was presented 13 times as a distractor in a set. The distractor icons were chosen at random and the location of these distractors in the array was randomized. The positions at which targets and distractors appeared in the array was carefully balanced across experimental trials. Participants in the equal experience condition were therefore given 60 simple, and 180 complex, icon trials (that is, a total of 240 experimental trials) at each time-of-day.

For both, equal and unequal conditions, the order in which the complex and simple displays were presented was counterbalanced so that half of the participants completed the complex displays first, while the other half completed the simple display first. For each participant the order of presentation of the trials was kept consistent, so for example participant 1 completed the simple display first and the complex display second at all test sessions.

3.3 Results

The percentage accuracies and the mean reaction times for correct responses, were analysed. The data was divided into two response types for the correct responses given; correct responses where the icon was present and correct responses where the icon was absent. Trials where the participant had made no response were not included in the analysis.

3.3.1. Response Times

Table 2 shows that performance varied according to the time-of-day where reaction times were fastest at 0900 in the unequal experience condition for present and absent complex icons and fastest at 1200 for present and absent simple icons. For the equal experience condition, the pattern was a little more variable, with fastest reaction times at 1200 for present complex icons, while absent complex icons were fastest at 0900. For present and absent simple icons in the equal experience condition fastest reaction times were at 2100. It can also be seen that reaction times were faster when participants had unequal experience than when they had equal experience. Also reaction times were slower when responding to icons that were absent from the display compared to when the icons were present in the display. Additionally, simple icons generated faster reaction times than complex icons.

A four-factor repeated measures analysis of variance was carried out to examine this data further. The factors were time of day (0900, 1200, 1500, 1800, 2100), icon complexity (complex versus simple), presence (icon present versus icon absent) and condition (unequal experience versus equal experience). See Appendix 1 for a full summary of results. The effect of time-of-day was significant ($F(4, 192) = 3.39$, $p<0.05$; see Figure 8). Response times were faster for simple, rather than complex, icons ($F(1,48) = 477.10$, $p< 0.001$). Responses were also faster when the icons were present in, rather than absent from, the display ($F(1, 48) = 480.51$, $p<0.001$). The difference in experience between experimental conditions did not significantly affect performance ($F(1,48) = 1.97$, p = ns; see Appendix 1(a) for details of this analysis).

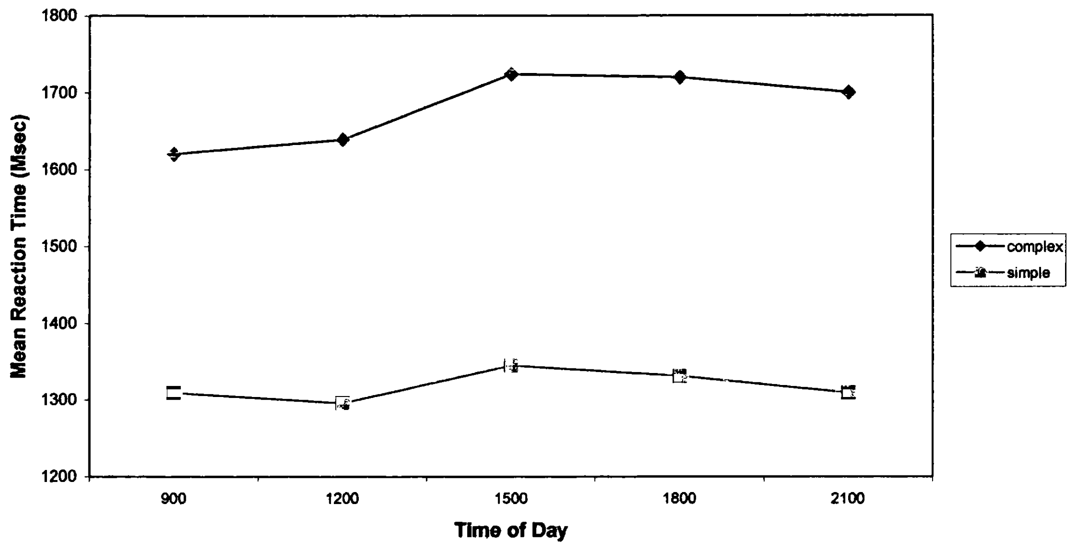
A significant interaction was observed between time-of-day and icon complexity ($F(4, 192) = 3.76$, $p<0.01$; see Figure 8) on response times. To investigate this further simple main effects were carried out which revealed significant differences in response times across the day for complex icons ($F(4, 192) = 4.19$, $p<0.01$) but not for simple icons ($F(4, 192) = 1.69$, p = ns) (see Appendix 1(b)). In order to explore the time-of-day effects existing in the use of complex icons further, Newman-Keuls analyses were carried out to explore exactly where significant differences lay. These analyses revealed that there were significant differences in

TOD	*Presence of Icon (P vs A), Icon Complexity (C vs S) & Condition(Unequal Experience vs Equal Experience)*							
	***P/C Unequal Experience**	**P/S Unequal Experience**	**P/C Equal Experience**	**P/S Equal Experience**	**A/C Unequal Experience**	**A/S Unequal Experience**	**A/C Equal Experience**	**A/S Equal Experience**
0900	1410.14 (276.85)	1059.11 (188.30)	1511.25 (255.76)	1193.22 (165.49)	1737.58 (391.71)	1413.66 (294.77)	1823.47 (289.85)	1572.82 (258.35)
1200	1440.50 (267.60)	1054.46 (174.04)	1504.86 (262.01)	1169.99 (174.33)	1753.44 (339.82)	1413.12 (278.81)	1857.44 (299.24)	1548.06 (224.38)
1500	1462.27 (401.72)	1108.83 (206.95)	1603.40 (298.17)	1242.58 (228.50)	1880.71 (411.07)	1484.25 (291.30)	1950.20 (323.06)	1544.51 (240.86)
1800	1515.42 (302.34)	1109.74 (206.93)	1596.22 (261.61)	1204.71 (204.93)	1831.65 (386.65)	1468.76 (298.29)	1935.70 (336.28)	1541.84 (245.04)
2100	1508.78 (300.40)	1084.56 (187.14)	1538.28 (210.02)	1169.57 (153.79)	1870.15 (375.67)	1483.31 (317.19)	1883.26 (243.51)	1500.43 (198.86)

Table 2: Mean (& standard deviations) Response Times for Present and Absent, Complex and Simple Icons for Each Condition at Each Time of Day

*P/C = present, complex; P/S = present, simple; A/C = absent, complex; A/S = absent, simple response times between 0900 and 1500, 0900 and 1800, 0900 and 2100 and also between 1200 and 1500, 1200 and 1800. Overall, this suggests that time-of-day effects are most marked between early/mid- morning (0900/1200) and the rest of the day (1500/1800/2100) for complex icons.

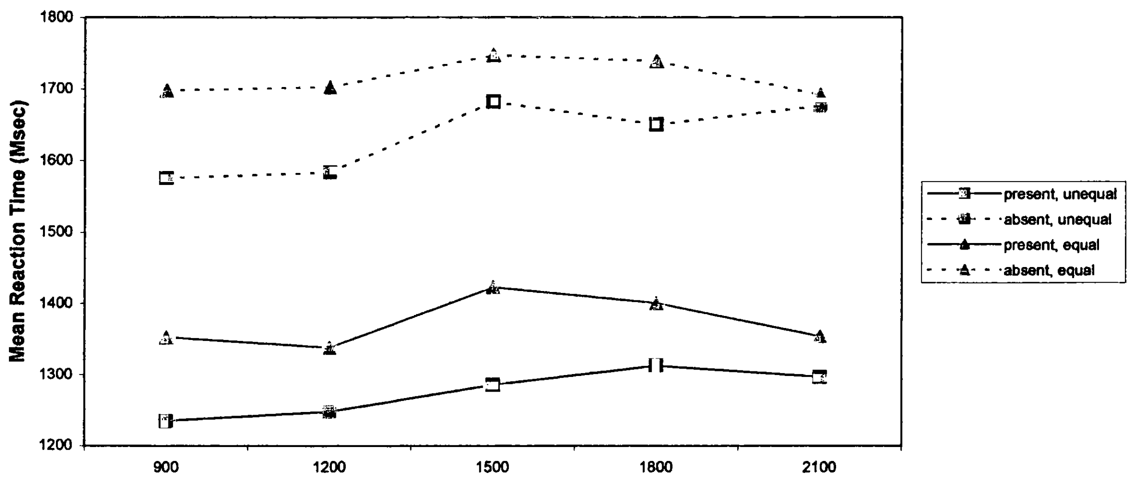
Figure 8: Mean Reaction Times for Complex and Simple Icons at Each Time of Day



Another interaction was observed between time-of-day, presence and condition ($F(4, 192) = 3.00$, $p<0.05$). From Figure 9 it can be seen that the response times for absent icons were much slower than response times for present icons, and that response times for the equal experience condition were slower than those for the unequal experience condition when the icons were both present and absent. However, this difference between the response times for each condition appeared to become less towards the end of the day. Simple interaction effects and simple simple main effects were carried out to investigate this further. Simple interaction effects revealed a significant interaction between time-of-day and presence for the unequal experience condition only ($F(4,192) = 2.88$, $p<0.05$), indeed the trends in Figure 9 for present and absent trials in the unequal experience condition suggest that the time-of-day trend is only significant for absent icons in the unequal experience condition. Accordingly, simple simple main effects revealed a significant time-of-day effect only for absent icons in the unequal experience condition ($F(4,192) = 3.78$, $p<0.01$). Further simple

simple main effects showed a trend for a difference between response times in the unequal experience and equal experience condition for present icons at 0900 ($F(1,48) = 3.76$, $p<0.058$) and at 1500 ($F(1,48) = 3.80$, $p<0.057$) but these did not quite reach significance (see Appendix 1(b) for details of this analysis). In order to explore the time-of-day effects for the absent icons in the unequal experience condition further, Newman-Keuls analyses were carried out to explore exactly where significant differences lay. These analyses revealed that there were significant differences in response time between 0900 and 1500 and between 0900 and 2100, again suggesting that it is the difference in response times between the early morning and the rest of the day that are important. No other interactions were significant.

Figure 9: Interaction Between Time of Day, Presence of Icon and Condition (Unequal versus Equal Experience)



3.3.2. Accuracy

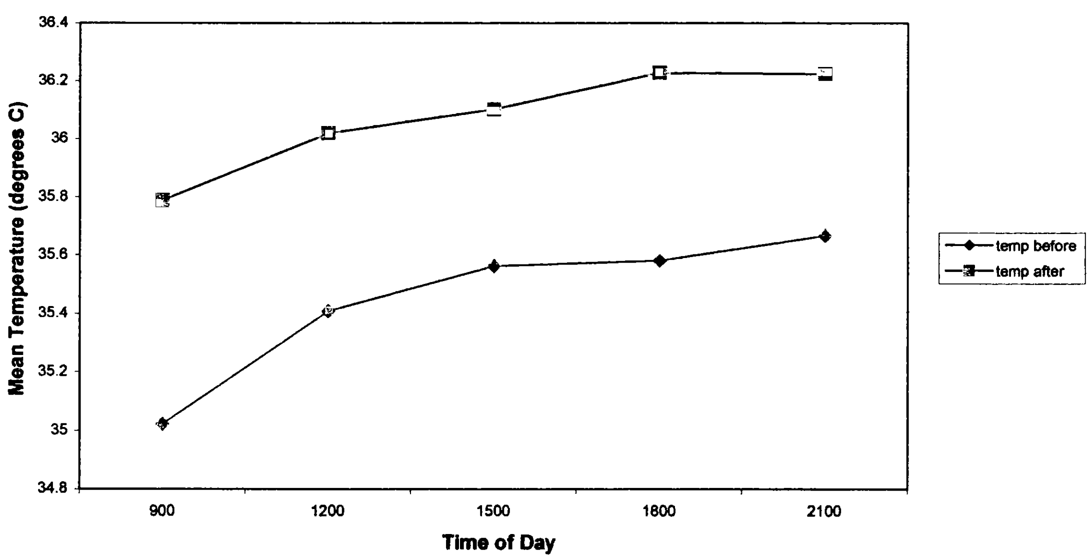
Due to the nature of the task involved, there was very little variation in the accuracy data obtained. For full details of percentage accuracy rates see Appendix 1(c). Mean percentage accuracies for each participant ranged from 84.78% to 99.00% with the mode being 96.56%. Mean percentage accuracies for each time-of-day ranged from 93.73% at 1500 to 94.51% at 0900 for the complex icons and from 95.19% at 1500 to 96.23% at 0900 for the simple icons. Overall the mean percentage was 95.69%. Due to the presence of these potential ceiling effects no further analyses were conducted.

3.3.3. Temperature

A three-factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900,1200,1500,1800,2100), for both the unequal experience condition and the equal experience condition. Significant differences in temperature values were found as a result of the time-of-day ($F(4,45) = 27.87$, $p<0.001$; see Figure 10). There was a significant difference between the temperature readings before and after each testing session at each time-of-day ($F(1,48) = 115.60$, $p<0.001$). There was no significant difference in temperature values as a result of differences in experience ($F(1,48) = 0.36$, p = ns). No interactions were observed (see Appendix 1(a) for full details of this analysis).

From Figure 10 it can be seen that temperature gradually increased over the course of the waking day with the lowest temperatures being seen at about 0900 and the highest at about 2100. The increase in temperatures after each test session can also be seen. To explore the time-of-day variations in temperature further, Newman-Keuls analyses were carried out to examine exactly where significant differences lay. These analyses revealed that temperature values differed significantly between 0900 and the rest of the day, 1200 and the rest of the day and also between 1500 and 2100.

Figure 10: Mean Temperature Values Before and After Each Time of Day



Comparison of the time-of-day trend seen for performance with that seen for temperature (Figures 8 and 10) shows no clear parallelism between the two.

Furthermore, comparison of the Newman-Keuls analyses for temperature and performance shows some significant differences in temperature that are not mirrored in performance.

3.4. Discussion

To summarise, response time varied in accordance with time-of-day when complex icons were presented. Response times were significantly faster at 0900 and at 1200 than at 1500 and 1800. Response times did not significantly differ between equal and unequal experimental conditions suggesting that user experience was not important in the tasks used. Responses were faster for simple, rather than complex, icons and when the target icons were present among, rather than absent from, the distractor array. Participants' temperatures significantly varied according to the time-of-day with temperature increasing over the course of the day from a minimum at 0900 to a maximum at 1800 but there was no obvious relationship between temperature and performance.

3.4.1. Time-of-Day Trends and Previous Research

The aim of Experiment 1 was to determine whether the time-of-day influences our ability to complete an icon task. This experiment also examined potential differences in the pattern observed between complex and simple icons. A significant time-of-day effect was observed and this was influenced by icon type. This experiment also examined the extent to which different types of icon influenced response times. Consistent with previous research, simple icons resulted in faster responses than complex ones probably because they could be discriminated using fewer features (Byrne 1993).

It is interesting that the diurnal trend was found to be significant only for the complex icons. Perhaps an explanation here, in terms of arousal, would be that performance on a task using more complex material is more dependent on the underlying arousal rhythm where lower levels of arousal are needed to maintain good performance, while performance on tasks using simpler stimuli is independent of the underlying arousal rhythm as the task is easy enough to be completed regardless of underlying arousal fluctuations (or indeed, any other mechanisms believed to underlie performance rhythms). Thus it seems only more complex icon tasks may be subject to diurnal variations in performance, perhaps because these tasks are more vulnerable to underlying arousal fluctuations.

It is interesting to note a slight difference concerning the timing of peak performance for each icon type; complex icons revealed fastest responses at 0900, while simple icons showed fastest responses at 1200. However, it is important to realise that the simple icons did not reveal a significant time-of-day effect and the difference in performance between 0900 and 1200 was not significant. Consequently, no conclusions can be drawn from this at present. Nonetheless, it is noteworthy that this slight difference in the timing of peak performance is consistent with previous memory load and working memory literature. To expand, for the reasons outlined in the introduction, the simple icons represented a lower memory load task, while the complex icons represented a higher memory load task. Working memory literature has found performance on working memory tasks to peak at about 1200 (Laird, 1925; Owens et al, 2000; Folkard, 1975), falling between the decreasing time-of-day trend seen for immediate/short-term *memory* and the increasing time-of-day trend seen for simple immediate *processing* tasks such as visual search (Folkard, 1983). Although the time-of-day effect was not significant for simple icons, the performance peak at 1200 seen for these icons is consistent with the above literature. Further, Folkard et al (1976) found the peak in working memory performance to occur earlier on with more highly loaded memory tasks and this is consistent with the 0900 peak in performance for more complex material. Therefore these results support the notion that the exact timing of the peak in performance depends on the memory load of the task, with the greater the working memory load a task requires the earlier in the day the peak in performance occurs (Folkard, 1983). Furthermore, the fact that the trends in Experiment 1 show a difference in the timing of peak performance according to the complexity of the icons, supports the concept that task demands are of great importance in determining the exact nature of the time-of-day trend observed (Smith, 1992).

A significant drop in performance at 1500 was observed for complex stimuli, however no significant improvement in performance was seen after 1500. Consequently, no support is provided for the well-documented post-lunch dip (for example, Kleitman, 1939,1963; Owens et al, 2000; Lenne et al, 1997).

Diurnal Trends in Performance and the Arousal Model and Other Mechanisms

The research reported here which showed performance on the more complex icon (high memory load) task to be slightly better than performance on the simple icon (low memory load) task at 0900 provides some, albeit weak, support for the assertion of the arousal theory that performance on complex, high memory load tasks would be best when arousal is low. However, no obvious parallelism between the temperature and performance trends was seen, failing to support the notion from the arousal model that the two are related.

We have seen that the arousal theory does in fact fail to provide a complete account of the mechanisms underlying diurnal variations in performance efficiency. Thus, perhaps more contemporary theories (for example, Monk et al, 1983; Monk et al, 1989; Dijk et al, 1992; Johnson et al, 1992; Folkard and Akerstedt, 1992) may be more accurate. However, as Carrier and Monk (2000) stated, dissecting the individual effects of the controlling processes on performance efficiency is not straightforward, especially considering that the resultant diurnal performance trends vary according to changes in task demands and/or individual differences.

Experiment 1 and the McFadden and Tepas (1997) Study

The results of McFadden and Tepas' (1997) study were reversed for the reasons outlined in section 3.1, so that the fastest response times are at 1200 for both participants in condition B (low memory load) and at 0900 in condition A (high memory load) for the second participant. The results of Experiment 1 therefore support McFadden and Tepas (1997). Further, in Experiment 1 significant main effects of time-of-day and complexity were found and there was a significant interaction between these two factors. Thus, in line with McFadden and Tepas' (1997) study: response times varied significantly according to the complexity of the icons and the time-of-day and there was a significant interaction effect with the time-of-day trend only being significant for complex icons in Experiment 1.

McFadden and Tepas' (1997) results were reversed because they believed that users of the original high memory load condition were limiting their search to the arrow component of the icons, thus converting this into a low memory load task. However, the icons used in the present study also contained an arrow that may allow users to limit their search (see Figure 6), yet the findings here are clearly consistent with previous literature. This raises the question about why this was not the case in McFadden and Tepas' (1997) study *before* the results were reversed. One possibility is differences in the stimuli used. This study used varying numbers of lines, that on the roads in Britain represent the distance to a motorway junction rather than 'North, South, East, West' as used by McFadden and Tepas (1997) and the route numbers used by McFadden and Tepas (1997) were replaced by shapes in this study. It seems likely that the difference between the icons presented in this study were not as obvious as that between those icons used by McFadden and Tepas (1997) therefore not allowing participants to strategically limit their search.

Although it seems that the icons designed for Experiment 1 may be advantageous in respect of the above, they also may have their limitations. Although each part of the icons designed for Experiment 1 can be seen on the roads in the UK in one form or another, they do not normally occur in combination. There may also be problems with familiarity differences with different parts of the sign being more familiar than others and also with more experienced drivers being more familiar with certain parts of the sign (for example, route) than less experienced drivers or non-drivers. Also the fourth distance sign and the rectangular route sign are not in existence on the roads. It is fair to conclude that the icons used would have been meaningless to participants, thus they do not have ecological validity unlike the stimuli used by McFadden and Tepas (1997). Indeed, one difference between this experiment and that of McFadden and Tepas (1997) that may be of importance, is that McFadden and Tepas (1997) used icons that would have been highly familiar to the participants (actual road signs were used), this was not the case in Experiment 1. However, in view of the consistency of the results of this experiment with those of McFadden and Tepas (1997) it seems unlikely that the lack of ecological validity exerted a great influence on the results.

Diurnal Performance Trends and Temperature

The temperature values were found to increase over the course of the day, increasing from its' lowest point at 0900 to it's maximum at 2100 confirming that participants demonstrated normal rhythms and supporting McFadden and Tepas (1997). No obvious relationship between temperature and performance was evident, lending support to Folkards' (1996) statement that performance on memory loaded tasks have rhythms that behave differently to those in body temperature.

Finally, temperature values after each test session were significantly higher than before each session. This can be explained in terms of increased cognitive activity raising the metabolic rate thereby increasing tympanic membrane temperature.

3.4.2. Experiment 2

Experiment 2 attempted to replicate the results found here for complex icons to prove the reliability of the trend observed. It seemed reasonable to focus on complex icons at this stage, as the time-of-day effect was only significant for this icon type. Both present and absent trials were tested again to examine whether one or both of these trials show diurnal performance fluctuations. Also it was interesting at this stage to consider the effect of combining each piece of information into one whole, rather than presenting each piece of information separately, to consider if the time-of-day effect remained and also to consider the effect of this on response times. It seems reasonable to suppose that if this information was presented as one whole performance may affected (see Figure 12, section 4.2.2.).

Although, in Experiment 1, one set of icons were considered to be more complex than the other while both sets were considered abstract in nature, the complexity and concreteness of icons was not measured, this was because our initial aim was to replicate McFadden and Tepas' (1997) study. In later experiments the effects of complexity and concreteness was considered more systematically by orthogonally varying these characteristics. The exact nature of the icons was potentially very important as we have already seen how variations in complexity, and therefore the exact task demands, can influence the resultant diurnal trend, thus it seemed fair to assume that more specific icon characteristics may exert similar effects. For instance, perhaps meaning is attached to icons differentially depending on the time-of-day, further if concrete icons eliminate time-of-day effects then the implications for cockpit design for example would be vital.

3.4.3. Conclusions

On the basis of these experimental findings, it can be concluded that pilots or workers who are monitoring chemical processing plants for example, will be subject to diurnal variations in their performance efficiency if the symbology that they are required to deal will in carrying out their work are complex. This appears to be because complex symbology places high demands on memory load. If simpler symbology, and thus a lighter memory load, is used the workers should be free from such performance variations.

Chapter 4

Experiment 2: The Effect of Multi-Feature versus Gestalt Icons

4.1 Introduction

Experiment 1 provided support for the notion that diurnal fluctuations exist in icon search tasks, effectively replicating McFadden and Tepas' (1997) earlier work. Interestingly however, Experiment 1 showed the diurnal performance trend to be significant for complex icons but not for simple icons, suggesting that the *type* of icon presented may determine the diurnal trends observed. Experiment 2 explored this possibility further by presenting two types of icon in which the features of an icon were either presented separately (multi-feature) or were presented to form a whole (gestalt). The complex (multi-feature) icons from Experiment 1 were used for the multi-feature condition (see Figure 11a) and the features were integrated to form a perceptual whole, or a gestalt, for the gestalt condition (see Figure 11b).

The gestalt or 'object-based' theories of visual attention propose that there is "a limit on the number of separate objects that can be perceived simultaneously" (Duncan, 1984 p.501). Neisser (1967) suggested that two stages exist in the object-based theory of visual processing, one where the visual field is divided into separate objects according to the gestalt principles of good continuation and proximity and a second where an object is perceived in greater detail. It was suggested that the first stage is parallel across simultaneously presented objects while the second stage is serial and thus it is this stage that imposes the limit on how many objects we can see at once. A great deal of support has been found for this object-based theory. For example, Treisman et al (1983) found that participants who had to identify a word and the location of a break in a box at the same time, performed the task more efficiently when the word was presented within the box rather than at the other side of the screen. Goldsmith (1998) also found the search process was most efficient when the features of the search were connected to the same perceptual object.

Duncan (1984) found that parallel preattentive processes divide the field into separate objects and focal attention, which follows, deals with one object at a time. Consequently, it seems reasonable to suppose that presenting each feature of the icons separately, as in the complex multi-feature icons in this experiment, is likely to encourage the features in each icon to be treated as separate objects. When focal attention is operating, serial processing of each feature is likely to slow processing time considerably. If however, these features were presented as a gestalt so that each icon is then perceived as one object, this should reduce the processing required and speed up responses.

The complexity of the gestalt and multi-feature icons was measured using Garcia et al's (1994) metric and were found to be equally complex. Therefore, on the basis of icon complexity research, these gestalt icons should be equally difficult to search for (see Byrne, 1993; McDougall et al, 2000). This effectively allows examination of whether gestalt icons alter the diurnal performance trend seen for multi-feature icons and whether gestalt icons improve processing speed.

In Experiment 1 an additional testing time (2100) was added to that originally done by McFadden and Tepas (1997). However, in these and subsequent experiments the decision was made to test only until 1800 since the results for Experiment 1 suggested that the significant differences in responses in general lay between 0900 and 1200 and the rest of the day and the 2100 time point did not show anything different from the 1800 time point.

Although the temperature trend failed to show any obvious relationship with performance in Experiment 1, temperature was again measured in this experiment, and throughout subsequent experiments, since it's relationship to performance trends may change with variations in task demands (see Campbell, 1992).

4.1.1. Aims

Experiment 2 examined the following:

1) Whether the observed trend in performance seen in Experiments 1 was reliable and could be replicated.
2) If the format in which the information contained in an icon was presented (multi-feature icons versus gestalt icons) influenced the observed diurnal performance trend.
3) If gestalt icons were advantageous in terms of performance (that is, response times and/or accuracy rates).
4) Whether the temperature trend would show a relationship to the performance trend seen for gestalt and/or multi-feature icons.

It was expected that the diurnal trend in performance observed for the complex multi-feature icons in Experiment 1 would be replicated. The memory load involved in Experiment 2 was high since the task involved: (a) a visual memory component; (b) a difficult 'yes'/ 'no' response; (c) icons that were difficult to discrminate. However, according to the object-based theories of visual attention, the gestalt icons should be perceived as one object rather than three separate objects (as is the case for multi-feature icons). On this basis it seemed likely that memory load would be less for gestalt, than for multi-feature, icons. On the basis of previous research (for example, Folkard et al, 1976; Folkard, 1983), the lower memory load in the gestalt condition was likely to produce a performance peak later in the day at 1200, while the higher memory load in the multi-feature condition was likely to produce peak performance at 0900. As gestalt icons encourage parallel search processes and reduce memory load in the way described above, it was expected that a decrease in response times would be observed for gestalt icons.

4.2 Method

The paradigm used in Experiment 2 was identical to that used in the complex icon trials in Experiment 1, with the exception that a new, gestalt, icon format was introduced.

4.2.1 Participants

Forty-eight undergraduates and postgraduates from the University of Wales Swansea were participants. Ten were male and thirty-eight were female. Six males completed the multi-feature condition while four completed the gestalt condition. Nineteen females completed the multi-feature condition while nineteen completed the gestalt condition. The age range of participants in the multi-feature condition was 18 years to 23 years; the mean age was 19 years 9 months (standard deviation, 1 year 9 months). The age range of participants in the gestalt condition was 18 years to 23 years; the mean age was 19 years 4 months (standard deviation, 1 year 6 months). Some obtained course credit of 3 hours for their participation while others received a payment of £10.

4.2.2. Materials & Apparatus

The materials and apparatus used for the multi-feature condition in Experiment 2 were identical to that used for the complex icon trials in Experiment 1 (see Figure 11a). The gestalt icon condition in Experiment 2 was identical to the above, with the exception that the features of the icons were presented in a gestalt format. To produce these icons each feature presented in the multi-feature icons were merged together (see Figure 11b).

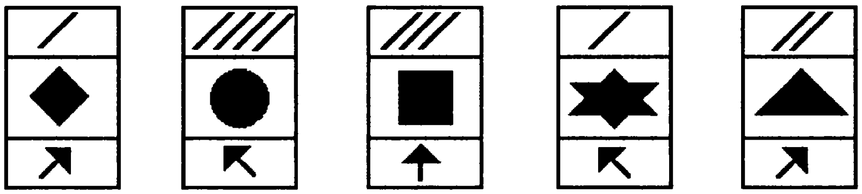


Figure 11 (a): Multi-Feature Icons Used for Experiment 2

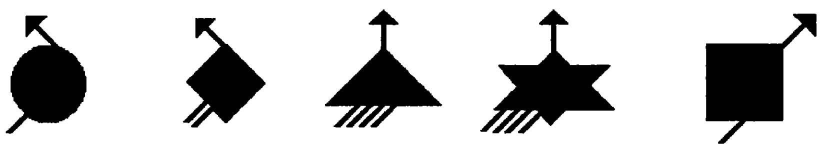


Figure 11 (b): Gestalt Icons Used for Experiment 2

The mean metric value (using Garcia et al's (1994) measure) for the icons used in the multi-feature condition was 5.50 (standard deviation, 1.29). The mean metric value for the icons used in the gestalt condition was identical, at 5.50 (standard deviation, 1.29).

4.2.3 Design

The participants were divided into 12 groups of four, one group of four participants were tested at a time. Each participant was tested once at each of the following times: 0900; 1200; 1500 and 1800. Each participant also completed a practice session at one of the test session times. The practice session a participant attended, and the order of testing thereafter, was counterbalanced using a Cyclic Latin Square (see Table 3) to ensure the administration of the experimental conditions was balanced across participants. Each participant's sessions were completed consecutively within 24 hours.

Group	*Practice*	*Test 1*	*Test 2*	*Test 3*	*Test 4*
1	0900	1200	1500	1800	0900
2	1200	1500	1800	0900	1200
3	1500	1800	0900	1200	1500
4	1800	0900	1200	1500	1800

Table 3: Cyclic Latin Square

4.2.4 Procedure

The procedure for Experiment 2 was identical to that for the complex icon trials in Experiment 1. Figure 12 shows an example of the interface display for the gestalt condition (Figure 7 in Chapter 3, section 3.2.4. shows a similar example for the multi-feature condition).

Figure 12 (a): Screen 1/Gestalt Condition, *target icons appears for two seconds.*

Figure 12 (b): Screen 2/Gestalt Condition, *target disappears & distractor array appears for six seconds, in this case target icon is present among the distractors. Participant should press the 'Q' key labelled with 'Y' for 'yes'.*

<u>Conditions, blocks of trials & randomisation of icons</u>

The presentation and randomisation of icons was identical to that for the complex icon trials in Experiment 1, except that only one block of 180 trials was presented to participants in each condition at each time point.

4.3 Results

The percentage accuracies and the mean reaction times for the correct responses were analysed. The data was divided into two response types for the correct responses given; correct responses where the icon was present and correct responses where the icon was absent. Trials where the participant had made no response were not included in the analysis.

4.3.1. Response Times

Table 4 illustrates a trend in performance that varied according to the time-of-day where reaction times were fastest at 0900 and slowest at 1500 for present and absent icons in the multi-feature condition and fastest at 1200 and slowest at 1500 for present and absent icons in the gestalt condition. It can also be seen that reaction times were faster in the gestalt condition than in the multi-feature condition. Further, reaction times were slower when responding to icons that were absent from the display compared to when the icons were present in the display.

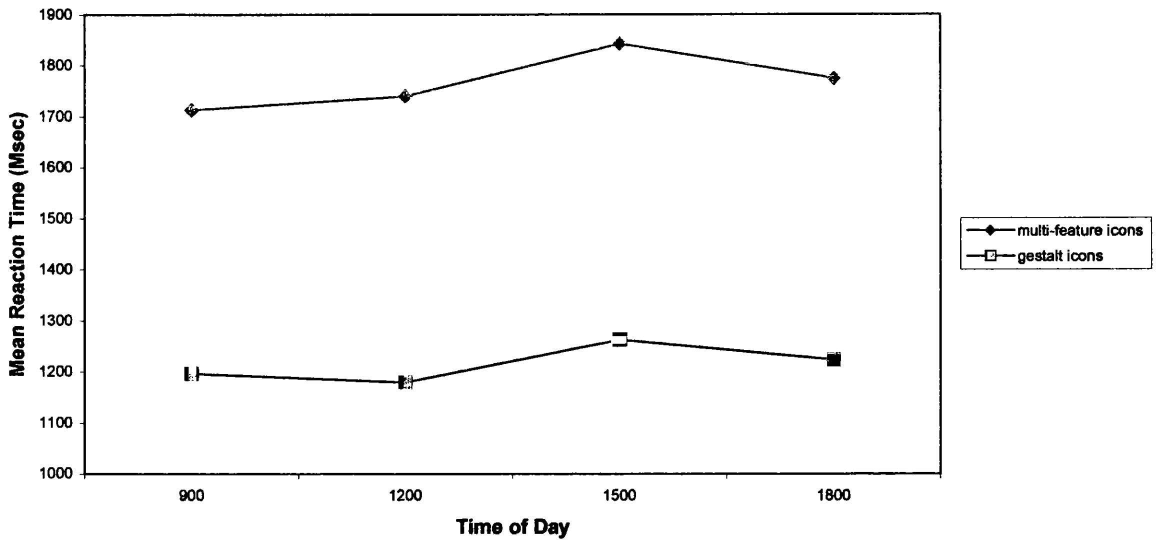
A three-factor repeated measures analysis of variance was carried out to examine the effects of time-of-day (0900, 1200, 1500, 1800), presence (icon present verses icon absent) and condition (multi-feature versus gestalt). See Appendix 2 for a full summary of results. The effect of time-of-day was significant ($F(3,138) = 6.02$, $p<0.001$; see Figure 13). Response times were fastest when the target icons were present among, rather than absent from, the display ($F(1,46) = 330.70$, $p<0.001$). There were also significant differences in response times between the multi-feature and gestalt conditions ($F(1,46) = 84.77$, $p<0.001$) with response times being strikingly faster for the gestalt icons (see Appendix 2(a) for details of this analysis).

From Figure 13 it can be seen that in addition to there being a marked difference in response times between gestalt and multi-feature icons, response times were fastest at either 0900 or 1200 with a dip in performance at 1500. In order to explore the time-of-day effect further, Newman-Keuls analyses were carried out to identify exactly where significant differences lay. These analyses revealed that there were significant differences in response times between 0900 and 1500, 1200 and 1500 and also between 1500 and 1800.

TOD	*Presence of Icon & Condition*			
	Present Multi-Feature Condition	**Present Gestalt Condition**	**Absent Multi-Feature Condition**	**Absent Gestalt Condition**
0900	1561.64 (254.56)	1077.47 (153.81)	1864.64 (302.86)	1313.99 (191.57)
1200	1582.25 (338.98)	1061.93 (122.66)	1897.36 (365.13)	1296.74 (152.12)
1500	1693.81 (260.84)	1157.20 (141.11)	1991.84 (304.62)	1368.60 (137.84)
1800	1623.50 (270.65)	1109.32 (181.57)	1926.38 (296.30)	1337.03 (228.13)

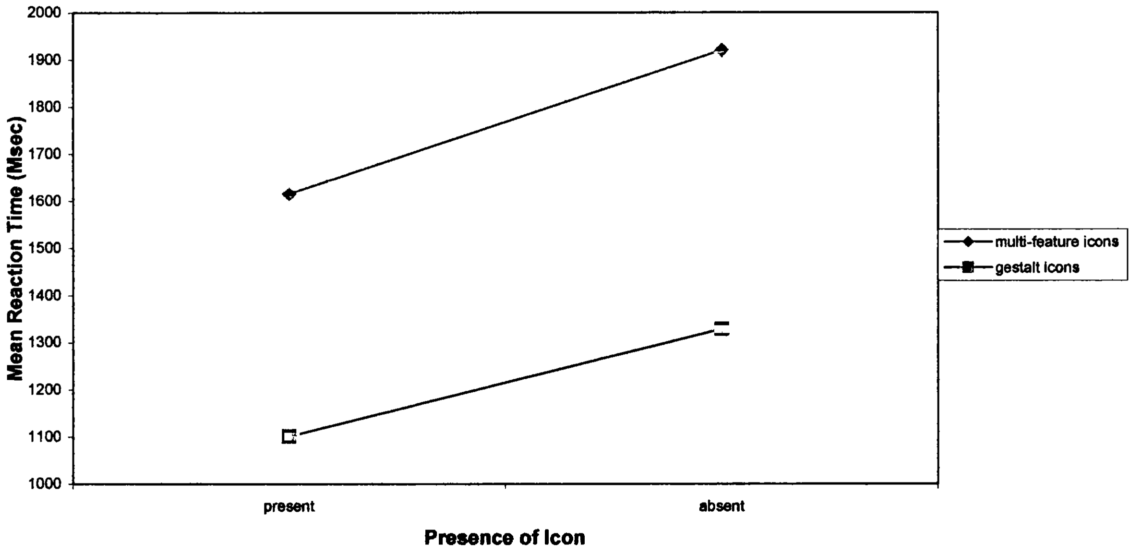
Table 4: Mean (& standard deviations) Response Times for Present and Absent Icons for Each Condition at Each Time of Day

Figure 13: Mean Reaction Times at Each Time of Day for Multi-Feature and Gestalt Icons



A significant interaction was observed between icon presence and multi-feature versus gestalt condition ($F(1,46) = 6.94$, $p<0.05$) on response times. Figure 14 clearly shows the difference in response times between multi-feature and gestalt icons, with response times for gestalt icons being faster than those for multi-feature icons. However the nature of any interaction is unclear with response times increasing for both icon types in absent trials. Indeed, simple main effects analyses showed the difference between the icon types was significant for both absent ($F(1,46) = 88.58$, $p<0.001$) and present trials ($F(1,46) = 67.73$, $p<0.001$; see Appendix 2(b) for details of this analysis). No other interactions were significant.

Figure 14: Interaction Between Presence and Icon Type



4.3.2. Accuracy

Due to the nature of the task involved in Experiment 2 there was very little variation in the accuracy data obtained. For full details of percentage accuracy rates see Appendix 2(c). Mean percentage accuracies for each participant ranged from 87.64% to 99.03% with the mode being 97.09%. Mean percentage accuracies for each time-of-day ranged from 94.27% at 1800 to 95.41% at 0900. Overall the mean percentage accuracy was 94.74%. As a result of these potential ceiling effects, no further analyses were carried out on this data.

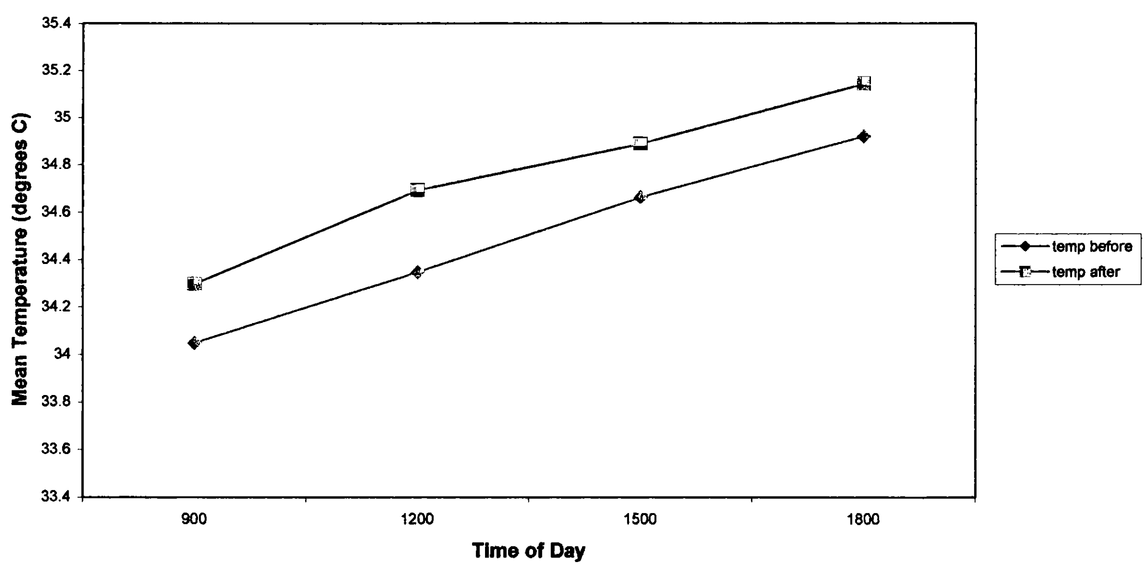
4.3.3. Temperature

A three-factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900,1200,1500,1800), for each icon type (multi-feature versus gestalt). Significant differences in temperature values were found as a result of the time-of-day ($F(3,138) = 116.24$, $p<0.001$; see Figure 15). There was a significant difference in temperature values before and after each testing session at each time-of-day ($F(1,46) = 124.21$, $p<0.001$). There was no significant difference in temperature values as a result of icon type ($F(1,46) = 1.52$, p = ns) (see Appendix 2(a) for details of this analysis).

From Figure 15 it can be seen that temperature gradually increased over the course of the waking day with the lowest temperatures being seen at 0900 and the highest at 1800. Figure 15 also shows the increase in temperatures after each test session. To explore the time-of-day variation in temperature further, Newman-Keuls analyses were carried out to explore exactly where significant differences lay. These analyses revealed that temperature values all differed significantly from one time-of-day to the next.

An interaction was observed between time-of-day and temperature before and after each test session ($F(3,138) = 4.06$, $p<0.01$). From Figure 15 it can be seen that the difference between mean temperatures before and after each test session was greatest at 1200 hours. However, simple main effects showed the difference between mean temperatures before and after each test session to be significant for all times of day: 0900 ($F(1,46) = 231.40$, $p<0.001$); 1200 ($F(1,46) = 305.75$, $p<0.001$); 1500 ($F(1,46) = 181.59$, $p<0.001$); 1800 ($F(1,46) = 217.27$, $p<0.001$) (see Appendix 2(b) for details of this analysis). No other interactions were significant.

Figure 15: Mean Temperature Values Before and After Each Time of Day



Comparison of the diurnal trend seen for performance with that seen for temperature (see Figures 13 and 15) illustrates that there is no obvious relationship between the two. Furthermore, comparison of the Newman-Keuls analyses for temperature and performance show that while there are consistent differences between each time-of-day for temperature this is not reflected in performance.

4.4. Discussion

Response times were subject to time-of-day effects. Response times were significantly different between 0900 and 1500, 1200 and 1500 and also between 1500 and 1800. Not surprisingly, response times were faster when the target icon was present among the distractor array. Temperature values significantly varied according to the time-of-day with temperature increasing over the course of the day from a minimum at 0900 to a maximum at 1800, but no relationship between temperature and performance was apparent.

4.4.1. Time-of-Day Trends and Previous Research

The aim of Experiment 2 was to replicate the results seen for the complex icon trials in Experiment 1 and to establish if the gestalt icon type was also subject to time-of-day effects. In both experiments response times were significantly faster at 0900 and 1200 than at 1500. However, only Experiment 2 showed a significant improvement in responses after 1500. Thus only Experiment 2, found evidence of a post-lunch dip in performance. Additionally, whereas in Experiment 1 icon type (complex versus simple) was found to interact with time-of-day, this was not the case in Experiment 2 despite the presence of a striking difference between response times for multi-feature versus gestalt icons.

Indeed, perhaps the most interesting finding was the marked decrease in response times seen for gestalt icons, theories of visual attention can offer an explanation for this. Object-based theories of visual attention propose two stages of processing; in the first, parallel processing is used and the visual field is divided into separate objects according to the gestalt principles (for example, good continuation, proximity) and in the second the object is seen in greater detail using serial processing and this limits the number of objects that can be seen at once (Neisser, 1967; Duncan, 1984, Goldsmith, 1998). Thus it is likely that the gestalt icons are perceived as one object during the first stage of processing and this reduces the limit imposed by the second stage with the whole object being seen at once. Consequently memory load is less for the gestalt icons and search times are reduced. Conversely, it is likely that the multi-feature icons encourage the icons to be divided into separate objects during the first stage, so that during the second stage each of these objects is considered one at a time thereby increasing memory load and search times.

Although gestalt icons were equally as complex as the multi-feature icons and so it could be argued that memory load was also equal, it seems the gestalt icons are *perceptually* different to the multi-feature icons and as such this serves to *reduce* memory load. Thus the perceptual difference between the multi-feature and gestalt icons makes the integrated object more effective. While in Experiment 1 it seemed that a reduction in complexity concurrently reduced memory load, Experiment 2 has shown that memory load can be reduced, without a reduction in complexity, simply by changing the format of the icons. It is important to note however that although the above theory is a plausible explanation of the results, these experiments were not designed to separate perceptual and memory load differences. Further, the difference in response times between gestalt and non-gestalt objects has never been shown before and it may take some time to establish the exact cause of this effect. Nevertheless, this finding is important in applied fields where speed of response is often critical and as such fully warrants further investigation.

Interestingly, as in Experiment 1, Experiment 2 shows a slight trend for the exact timing of peak performance to change according to icon type and concurrent changes in memory load. Here the multi-feature icons show a slight performance advantage at 0900 that is not seen for the gestalt icons. This could be explained in terms of differences in memory load between the two icon types (discussed above) and indeed this would be consistent with previous working memory literature that has found a peak in performance at 1200 (Laird, 1925; Owens et al, 2000; Folkard, 1975; Folkard, 1983; Folkard et al, 1976) or earlier on more highly loaded memory tasks (Folkard et al, 1976). So again it seems the exact task demands are determining the diurnal trend seen (Smith, 1992). However, as in Experiment 1, clearly the difference in performance at these times is not significant and no interaction between time-of-day and icon type was observed, as such no definite conclusions can be drawn at present, although it is nonetheless interesting to note the trend seen.

In Experiment 2 a dip in performance was found at 1500. It seems reasonable to conclude that this is attributable to the well-documented post-lunch dip (for example, Kleitman, 1939, 1963; Owens et al, 2000; Lenne et al, 1997). It is interesting to note that this afternoon dip in performance was not seen in Experiment 1. This can perhaps be explained by previous work that has shown the post-lunch dip to be pliable in that it can be made worse by ingestion of a high carbohydrate meal or that it can be removed by noise for instance (for example, Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b, 1990; Smith, 1988).

Response times in Experiment 2 were faster when the target icon was present in the display, this was also the case in Experiment 1. The most likely explanation is that the present icons were identified during stage 1 of the search when a parallel search of all icons took place, while absent trials would have resulted in a stage 2 search where the items in the display were searched serially, thus the latter would take more time to complete. Further, the difference in response times between the multi-feature and gestalt icons was greater when the target icons were absent from the array, this can most probably be explained in terms of the absence of the icon further increasing the difficulty of the task when using multi-feature icons thereby further increasing the difference in response times between the two icon types in the absent trials.

Diurnal Trends in Performance and the Arousal Model and Other Mechanisms

Again, the slight difference in the timing of the peak performance between more difficult (multi-feature icons) and less difficult (simple icons and gestalt icons) tasks can be taken as providing further support, albeit weak, for the arousal theory that postulated that the optimum level of arousal is high for low memory load tasks. However, the arousal framework also asserted that temperature should be related to performance, but no relationship between temperature and performance was apparent, supporting previous research (for example, Owens et al, 2000). The results therefore suggest that more contemporary theories may be more accurate when they propose that performance trends are governed by a homeostatic process and by a circadian timing system (for example, Monk et al, 1983; Monk et al, 1989; Dijk et al, 1992; Johnson et al, 1992; Folkard and Akerstedt, 1992). Support is also provided for Carrier and Monks' (2000) suggestion that performance rhythms are probably best considered independently of physiological rhythms (Carrier and Monk, 2000).

4.4.2. Experiments 3 and 4

Since the perceptual form of icons appeared to influence response times and to slightly affect diurnal performance trends, it may be that specific icon characteristics affect performance across the day. Also although the complexity of the icons was measured using Garcia et al's (1994) metric, the complexity/concreteness of the stimuli was not systematically varied because our aim here was to replicate the results seen for the complex set of trials in Experiment 1. Additionally, complexity had to be equated in order to compare gestalt verses multi-feature icons. Experiments 3 and 4 explored the role of icon characteristics in mediating time-of-day effects further, by examining the role of icon concreteness. Concreteness is thought to be important because research has shown that concrete icons minimise articulatory distance (Hutchins et al, 1986) and as a result, identification performance improves (Garcia et al, 1994). Experiments 3 and 4 orthogonally varied the concreteness and complexity of icons, where icons can be concrete or abstract independent of their level of complexity.

Experiment 4 further enhanced the everyday applicability of the tasks used by employing a procedure analogous to the process often undertaken by new interface users (for example, a learner driver) by requiring users to directly attach meaning to each icon. So far the effect of having to attach a meaning to an icon has not been considered. This is of importance when consideration is given to the fact that when searching for a road sign, for example the distance to your motorway junction, the driver knows the meaning of the sign he/she is looking for and must match the sign when he/she sees it to this meaning. In Experiment 4 function labels had to be matched to the icon in much the same way as the meaning must be matched to the sign on the roads. Although some of the icons in Experiment 4 will be quite obscure, they will be readily learned, analogous to the way a learner driver learns to match meaning to road signs. Thus the process of attaching meaning to a symbol and the influence of this task on the time-of-day trend was directly examined.

A factor that was not yet systematically considered was the *exact* demands of the *whole* testing procedure. The procedures used so far were actually quite complex, requiring the use of working/visual memory and a difficult 'yes'/ 'no' keyboard response type, where participants had to press one key if the icon was present among the display and a different key if it was not. Furthermore, the icons designed for use in the experiments thus far appeared to lack distinctiveness making them difficult to discriminate. It was possible that it was one or more of these task requirements that was responsible for the resulting time-of-day performance trend. It would therefore be useful to attempt to systematically vary these requirements to establish exactly what was important. Consequently, further experiments, beginning with Experiments 3 and 4, attempted to systematically vary the demands of the experimental procedures, to establish whether any of these factors fully or partly determine the exact time-of-day trend.

Thus Experiment 3 examined the diurnal trends for icons that were orthogonally varied in terms of their complexity and concreteness and appeared to be more distinctive and therefore easier to discriminate between. A simple search task paradigm that involved no visual memory and a simple mouse click response was used. Experiment 4 was identical to Experiment 3 with the exception that this experiment examined the effect of meaning.

4.4.3. Conclusions

It can be concluded that those working with symbology are equally susceptible to diurnal performance fluctuations whether the icons used form a perceptual whole, or a gestalt, or whether each piece of information to be communicated is presented separately within the icon. Through encouraging parallel search processes and reducing memory load, gestalt icons dramatically reduce response times. On the basis of these findings it seems advisable to design symbology using a gestalt format, especially for time-critical applications such as air-traffic control.

Chapter 5

Experiments 3 & 4: The Effect of Changing Task Demands

5.1 Introduction

Experiments 1 and 2 have provided good evidence that tasks that require the use of symbology are as susceptible to diurnal fluctuations in performance efficiency as other tasks. These experiments showed that the complexity of icons and the format in which they are presented, influence participants' response times, where simple and gestalt icons generated faster responses than complex and multi-feature icons. The results of Experiment 1 were consistent with previous research, which has shown that simple icons reduce visual search time (McDougall et al, 1996; Byrne, 1993). The type of icon presented also had a little influence over the time-of-day trends observed. Performance trends early in the day slightly varied according to changes in memory load resulting from manipulations of the complexity of the icon (see Experiment 1) or the icon format (gestalt versus multi-feature, see Experiment 2).

Experiments 3 and 4 attempted to extend the findings of Experiments 1 and 2 by varying both the complexity and concreteness of icons shown to participants. The icon set used had previously been employed in a study by McDougall et al (1998; see Figure 16). In their study, two types of task were used. The first was a *search task* where a target icon had to be identified with the same icon appearing among a distractor array. This task was used in Experiment 3. The second task was a *search and match task* where a function label had to be matched to the icon representing this function. So participants not only had to search for the icon on the screen, they had to match the icon to the function given. This task was used in Experiment 4. 'Gestalt' icons were used as these have been shown to be more beneficial in terms of speed of response (see Experiment 2). However, these icons were very different to those used in Experiment 2. According to Goldsmith's (1998) principles however the nature of the icons used are still likely produce the tendency for them to be perceived as coherent wholes and as such can be regarded as gestalts. It is important to note that the icons used by McDougall et al (1998) could be considered to be more distinctive than those used in Experiments 1 and 2. The icons in Experiments 1 and 2 had very similar features that made them difficult to discriminate from one another (see Figure 6) while the icons used in Experiments 3 and 4 were easier to discriminate (see Figure 16). Byrne (1993) believed that icons that are "simple and easily discriminable" (p.452) are advantageous to performance, whereas icons that are "complex and difficult to discriminate quickly" (p.452) are not advantageous. Likewise, Fisher et al (1992) stated that distinctive icons are easier to find quickly in displays in which they may be presented with other icons. Further, Magyar (1990) noted that distinctive icons are not as easily confused with other icons as non-distinctive icons may be. Indeed, comparison of the icons used in these experiments (see Figures 6 and 16) illustrates how the icons used in Experiments 1 and 2 could be easily confused with one another, while those in Experiments 3 and 4 may not be so easily confused, effectively highlighting the distinctiveness of this new set of icons. The end result is that distinctiveness may encourage a 'pop-out' effect of target stimuli due to its' uniqueness or dissimilarity to the environment surrounding it, however it has been found that practice is required for this pop-out effect to occur (Boersema and Zwaga, 1996). Nonetheless, this new set of icons effectively allowed a comparison of the effects of icon complexity and icon concreteness.

McDougall et al (1996) found that icon complexity had a strong influence on search tasks while icon concreteness, which is closely related to meaning, was more important in determining performance in tasks where meaning was involved (Paivio et al, 1968; Gilhooly and Logie, 1980). As a result, one might expect icon complexity to be of primary importance in Experiment 3, which involved only visual search. In contrast, one might expect icon concreteness will be of primary importance in Experiment 4 where meaning is vital to the completion of this semantic memory task. Furthermore, research has shown that although concreteness influences the initial understanding of an icons' meaning and therefore responses are initially much faster for concrete icons (McDougall et al, 2000), once meaning is learned these concreteness effects disappear (McDougall et al, 2001) and whether the icon is complex or simple begins to play a more important role (McDougall et al, 1998). Consequently, one might also expect complexity effects in Experiment 4 with complex icons producing slower response times even when the icon set has been learned (McDougall et al, 1996).

The previous experiments have highlighted the possibility that it may be the demands of the task that are producing the time-of-day trends observed, however previous experiments only examined this in terms of different icon types, the possibility that it may be the response type used for instance, has not yet been considered. Experiments 3 and 4 will allow consideration of the influence of changing task demands by using a simple search task (where no meaning had to be linked to an icon and no working memory was required) and a search and match task (where meaning had to be attached to an icon and semantic memory was accessed to do so). One might expect the performance trends of these tasks to follow those seen for previous experiments using tasks that have involved similar cognitive processes. For instance, Folkard and Hill (2002) stated that performance on simple serial search speed tasks, which involve little or no working memory component, has been found to peak in the evening. Millar et al (1980), Tilley and Warren (1983) and Smith (1987a) found that performance on a semantic memory task also improves later in the day. Further, Whitney and Williams (unpublished) found semantic access to be slowest mid-afternoon and attributed this to the post-lunch dip. However, it has been shown that when the nature of a semantic memory task is changed then the original diurnal performance trend may change. Indeed, Smith (1987a) found that if participants had to constantly use different retrieval strategies, the time-of-day trend disappeared altogether. The performance trend seen for semantic memory tasks can be explained by work showing that more maintenance processing tends to occur in the morning while more semantically based elaborative processing is used in the evening (for example, Folkard, 1979, 1980; Lorenzetti and Natale, 1996; Oakhill 1986a, 1986b, 1988; Marks and Folkard, 1988). Folkard (1983) related this to changes in arousal, stating that more attention may be given to the physical characteristics of the information in the morning when arousal is low, while more attention may be given to the meaning of the material later in the day when arousal is higher. Folkard and Hill (2002) explained the different diurnal trends shown by tasks using different types of memory by suggesting that each type of memory uses different cognitive subsystems that each have different circadian systems controlling them thereby resulting in different diurnal performance trends.

The task in Experiment 3 differed from previous tasks, in that the tasks used in Experiments 1 and 2 were quite difficult involving a high memory load requiring the use of: (a) visual memory; (b) a difficult 'yes'/ 'no' response; (c) non-distinctive icons that were difficult to discriminate. Thus there are many features of the task used in Experiments 1 and 2 that could be responsible for the observed time-of-day trend. Experiment 3, however, was a simple search task where the icon to be searched for did not disappear from screen and therefore did not need to be remembered thus no visual memory was required (see Figure 17). Furthermore, participants were required simply to click on the matching icon with the mouse, as a consequence of this aspect of the task there were no absent icons – all were present in the display to be searched. Finally, the icons used may have been easier to discriminate due to their relative distinctiveness. In this way it was hoped to test the occurrence of time-of-day effects in a simple icon task that used no working memory, so that, through comparisons with other experiments, the effects of the exact demands of the task could be considered.

Experiment 4 used the same paradigm as Experiment 3 but instead of a simple search task where icons were matched to icons, function labels were matched to icons, in this way participants would have to directly attach meaning to a symbol. Thus although no visual memory component was involved and an easy mouse click response was used, the icons did not simply need to be identified, but their meaning had to be matched to them, resulting in semantic memory involvement.

5.1.1. Aims

Experiment 3 examined the following:

1) Whether diurnal performance trend was observed in a simple icon search task.
2) Whether the diurnal trend observed varied according to icon type.

Experiment 4 examined the following:

3) Whether time-of-day effects existed in icon tasks where meaning was directly attached to an icon.
4) Whether the diurnal performance trend varied between the previous simple search task and this task where an additional semantic memory component was added.

Both Experiments 3 and 4 examined:

5) Whether the temperature trend showed a relationship to the performance trend seen for either experiment.
6) Both experiments also examined the role of icon concreteness and icon complexity in determining performance.

It was expected that a significant diurnal trend in performance would again be seen for both experiments. In accordance with previous research outlined above, in Experiment 3 the performance trend seen was expected to show a peak in the evening consistent with research into tasks using no working memory, while the trend for Experiment 4 was also expected to show improved performance later in the day. Simple icons were expected to show faster response times than complex icons.

Finally, in Experiment 4 where meaning was important, concrete icons were expected to be interpreted more readily than the abstract ones as they bear a closer resemblance to the object of representation, however, after a few trials this concreteness effect was expected to disappear and whether the icon was complex or simple was expected to play a more important role. Whether the performance trend would show a relationship to the temperature trend could not be predicted, although this seemed unlikely given the findings in Experiments 1 and 2.

Experiment 3:Time-of-day effects in a simple search task

Experiment 3 was a simple search task using gestalt icons, which varied in their complexity and concreteness. Only a simple mouse click response was required and there was no visual memory component to this task.

5.2 Method

Experiment 3 used a visual search paradigm where a target icon was presented on screen and this target was then matched to the same icon appearing among an array of 9 icons that were subsequently shown on screen.

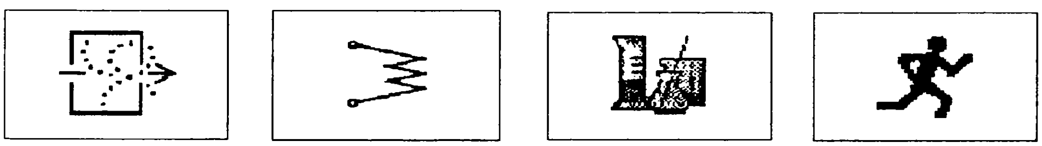
5.2.1. Participants

Twenty-four undergraduates and postgraduates from the University of Wales Swansea participated, nine were male and fifteen were female. The age range of participants in Experiment 3 was 18 to 25 years; the mean age was 20 years 1 month (standard deviation, 2 years 3 months). Some received course credit while others received a payment of £10.

5.2.2. Materials & Apparatus

Icons were presented using a self-paced computer program that moved on to the next trial as soon as the participant had responded and automatically recorded reaction times and accuracy rates. The visual search task was presented on Pentium 166 MHz computers. The screen settings on all computers were set to 1024x768 pixels. Reaction times were measured using the systems' multimedia timer, allowing measurement to within a 1millisecond resolution. A standard Microsoft mouse was used with each computer, with the track speed set halfway between slow and fast. The use of a mouse as an input device gives an estimated uncertainty of up to $\pm$ 30 milliseconds (see McDougall and de Bruijn 1999b). A Braun Thermoscan thermometer (Model IRT 3520) was used to measure participants' temperatures before and after each test session.

All icons were black, white and grey. Seventy-two icons were used, there were eighteen icons in each icon category of: abstract-complex; abstract-simple; concrete-complex; concrete-simple (see Figure 16 for example). Each icon was representative of something and function labels indicating what each icon represented can be seen in Figure 21. McDougall et al (2000) orthogonally varied the concreteness and complexity of the icons using ratings of each characteristic on a 1-5 scale (a rating of 5 indicated that the icon was definitely concrete while a rating of 1 indicated that the icon was definitely abstract). Statistical analyses revealed the ratings differed in accordance with the classification of each icon type. Concrete-complex and concrete-simple icons had higher ratings than abstract-complex or abstract-simple icons. Likewise, concrete-complex and abstract-complex icons had higher complexity ratings than concrete-simple and abstract-simple icons. Thus concreteness and complexity had been orthogonally varied in McDougall et al's (2000) experiment.



Abstract-Complex *Abstract-Simple* *Concrete-Complex* *Concrete-Simple*

Figure 16: Examples of Each Icon Type Used in Experiment 3

5.2.3. Design

Forty-eight participants were used in total for Experiments 3 & 4, which were conducted simultaneously. Each participant was allocated, on an alternating basis, to a particular experiment. The participants were divided into twelve groups of four, one group of four participants were tested at a time. Each participant was tested once at each of the following times: 0900; 1200; 1500; 1800. Each participant also completed a practice session at one of these times. The practice session a participant attended, and the order of testing thereafter, was counterbalanced using a Cyclic Latin Square (see Table 5) to ensure the administration of the experimental conditions was balanced across participants. Each participant's sessions were completed consecutively within 24 hours.

Group	*Practice*	*Test 1*	*Test 2*	*Test 3*	*Test 4*
1	1800	0900	1200	1500	1800
2	0900	1200	1500	1800	0900
3	1200	1500	1800	0900	1200
4	1500	1800	0900	1200	1500

Table 5: Cyclic Latin Square

5.2.4. Procedure

To start each experimental trial, participants were required to click on an "ok" button present on the screen using the mouse. Once participants had clicked on the "ok" button it turned grey and the target icon was presented on screen for two seconds, after which the "ok" button turned black and participants were required to click on it once more. This second click on the "ok" button ensured that the starting position of the mouse was always at the same position on the screen. Once this had been done, the array of 9 icons appeared on the screen from which the participants were required to select the target icon as quickly as possible using the mouse and timing began. Participants were instructed to respond as quickly and as accurately as they could at all times. Timing ceased when an icon was selected. The next trial would then begin and the whole procedure was repeated. No feedback was given regarding the correctness of response. If no response was given within 6 seconds, the program moved on to the next target automatically (see Figure 17). This was repeated for each of the three blocks. There was a two-minute break between each block.

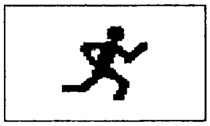


Figure 17 (a): Screen 1/Experiment 3, *target icon appears for two seconds before the distractor array appears.*

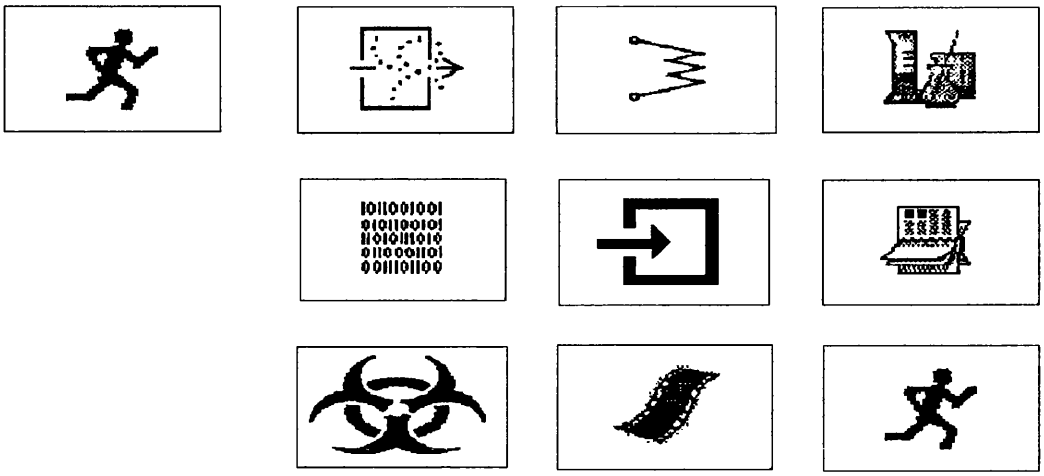


Figure 17 (b): Screen 2/Experiment 3, *target remains on screen as the distractor array appears, the participant should click on the matching icon using the mouse.*

Eardrum temperature readings were taken before and after each test session and were taken from the right ear on every occasion. The procedure was identical at every test session.

Conditions, blocks of trials and randomisation of icons

Presentation of the targets and the distractor arrays was randomised so that each had an equal chance of being sampled and the participants saw a different set of stimuli during each test session. In each session there were three blocks of 72 trials. Each block of 72 trials used the same set of 72 icons. Each icon was presented once as a target in each block of 72 trials. Each icon appeared 8 times as a distractor in each block of 72 trials. The array of 9 icons consisted of the target icon and 8 distractors. The 8 distractor icons in each trial consisted of 2 icons from each of the four experimental conditions (abstract-complex; abstract-simple; concrete-complex; concrete-simple). The location of these distractors within the array was randomised. Target icons from each of the icon conditions were presented twice at each of the 9 possible positions on the display. Each position (1-9) on the display was used 8 times.

5.3 Results and Discussion

The percentage accuracies and the mean reaction times for the correct responses were analysed. Trials where the participant had made no response were not included in the analysis.

5.3.1. Response Times

Table 6 shows a trend in performance that varied according to the time-of-day where generally reaction times were fastest at 1800 and slowest at 0900. Performance generally improved between 0900 and 1200 before dropping again at 1500. It is also shown that generally reaction times were fastest in the simple condition. Also, abstract icons generally produced faster reaction times than concrete icons when the icons were also simple and conversely, concrete icons were faster than abstract when the icons were also complex.

A four-factor repeated measures analysis of variance was carried out to examine the effects of the time-of-day (0900, 1200, 1500, 1800), icon concreteness (abstract versus concrete), icon complexity (simple versus complex) and block number (1, 2, 3). See Appendix 3 for a full summary of results. The effect of the time-of-day was significant ($F(3,69) = 4.01$, $p<0.05$; see Figure 18). Responses were fastest when icons were simple rather than complex ($F(1,23) = 121.97$, $p<0.001$). Although there was a trend for a variation in performance over blocks of trials, with response times generally becoming faster over the 3 blocks of trials at each time-of-day, this did not quite reach significance ($F(2,46) = 3.02$, $p = 0.059$). Icon concreteness did not significantly affect performance ($F(1,23) = 0.18$, $p =$ ns) (see Appendix 3(a) for details of this analysis).

From Figure 18 it can be seen that all icon types revealed similar diurnal performance trends, where generally response times were slowest at 0900, fastest at 1200 or 1800, showing a slight increase in response times at 1500. In order to explore the time-of-day trends seen for each icon type further, Newman-Keuls analyses were carried out to examine exactly where significant differences lay. These analyses revealed that overall there were significant differences between 0900 and 1200 and between 0900 and 1800. This suggests that time-of-day effects are most marked between early morning (0900) and later in the day (1200/1800). It is interesting to note that, in contrast to Experiments 1 and 2, a significant difference between 0900 and 1200 was observed.

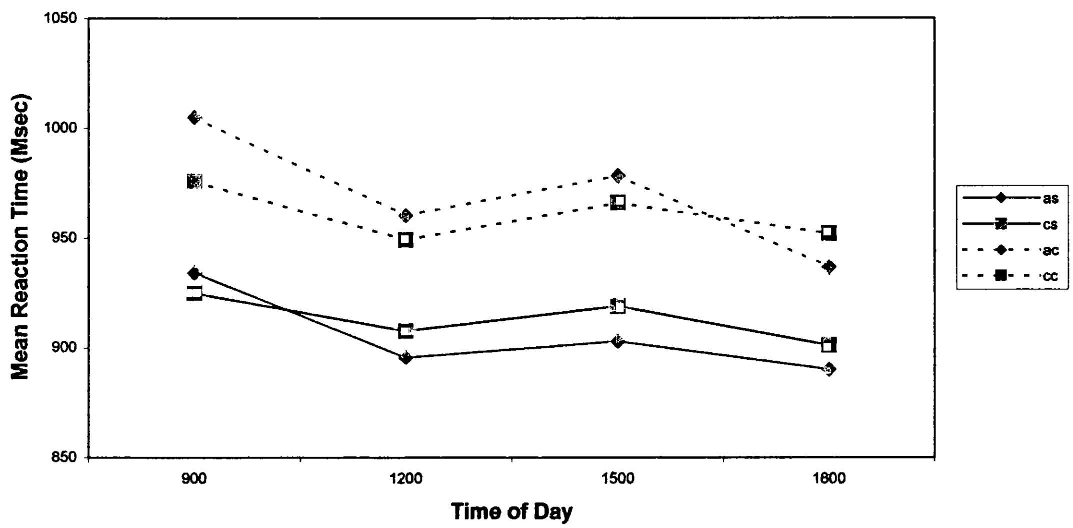
TOD	*Condition & Block Number*											
	***ASb1**	**ASb2**	**ASb3**	**ACb1**	**ACb2**	**ACb3**	**CSb1**	**CSb2**	**CSb3**	**CCb1**	**CCb2**	**CCb3**
0900	948.45 (101.19)	937.48 (162.56)	915.24 (119.49)	1027.50 (116.72)	984.29 (148.37)	1002.68 (166.71)	931.17 (132.14)	924.60 (95.69)	918.64 (107.35)	991.06 (110.82)	991.91 (125.64)	944.87 (135.20)
1200	909.62 (136.10)	889.79 (122.50)	887.13 (120.03)	968.19 (128.31)	948.92 (157.43)	963.99 (136.96)	917.68 (135.18)	912.01 (148.50)	894.04 (122.89)	926.68 (227.83)	954.45 (145.23)	927.67 (133.70)
1500	913.84 (114.70)	892.04 (97.78)	901.97 (127.06)	983.62 (136.79)	963.38 (116.17)	988.87 (149.63)	925.39 (127.16)	923.51 (115.55)	907.56 (117.88)	971.59 (147.79)	959.58 (152.39)	966.75 (137.98)
1800	903.56 (124.07)	883.07 (113.12)	882.90 (111.61)	929.18 (113.56)	950.56 (121.06)	931.06 (122.91)	911.52 (117.71)	893.06 (131.17)	899.04 (123.47)	949.64 (117.80)	964.47 (141.89)	941.57 (118.71)

Table 6: Mean (& standard deviations) Response Times for Each Condition at Each Time-of-Day

*AS = abstract-simple; AC = abstract-complex; CS = concrete-simple; CC = concrete-complex

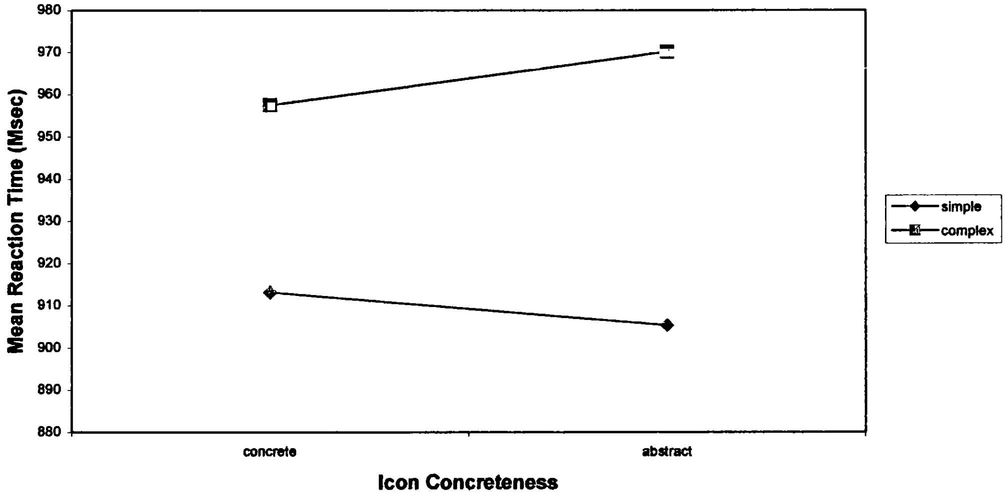
(note: these abbreviations also apply to figures where reference is made to each icon type, throughout Experiments 3-6)

Figure 18: Mean Reaction Times for Each Icon Type at Each Time of Day



A significant interaction was observed between icon concreteness and icon complexity ($F(1,23) = 5.04$, $p<0.05$) on response times. Figure 19 shows response times were fastest when simple abstract icons were shown and were slowest when abstract complex icons were shown. From Figure 19 it can be seen that the difference in response times between simple and complex icons became greater when the icons were abstract. Nevertheless, simple main effects revealed the difference in response times between simple and complex icons was significant when the icons were both abstract ($F(1,23) = 123.38$, $p<0.001$) and concrete ($F(1,23) = 34.96$, $p<0.001$; see Appendix 3(b) for details of this analysis). No other interactions were significant.

Figure 19: Interaction Between Icon Concreteness and Icon Complexity



5.3.2. Accuracy

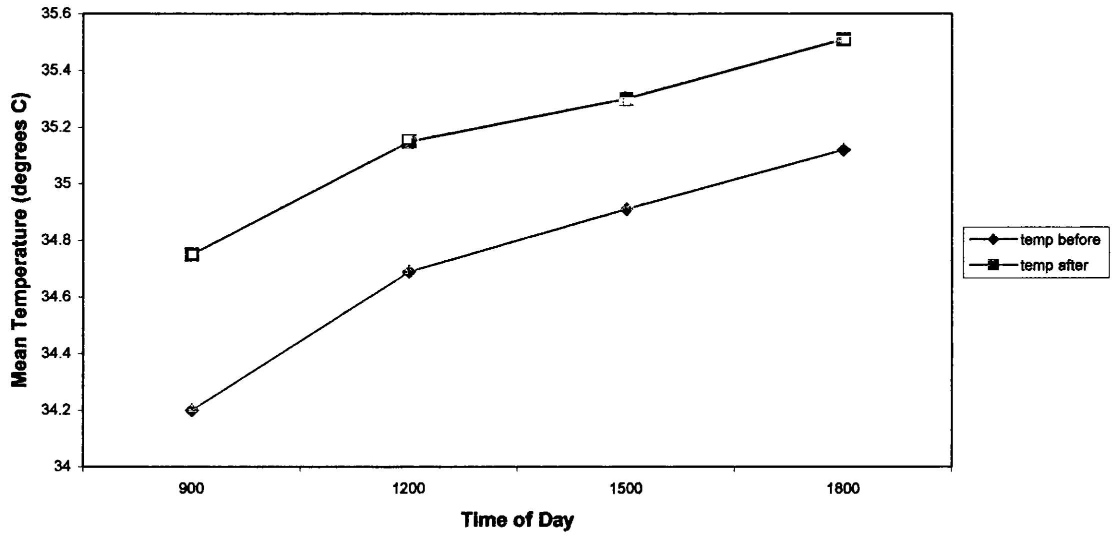
Due to the nature of the task involved in Experiment 3 there was very little variation in the accuracy data obtained. For full details of percentage accuracy rates see Appendix 3(c). Mean percentage accuracies for each participant ranged from 93.98% to 100% with the mode being 100%. Mean percentage accuracies for each time-of-day ranged from 99.48% at 1800 to 99.81% at 0900. Overall the mean percentage accuracy was 99.64%. Due to the presence of these ceiling effects no further analyses were conducted.

5.3.3. Temperature

A two-factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900, 1200, 1500, 1800). Significant differences in temperature values were found as a result of the time-of-day ($F(3,69) = 42.79$, $p<0.001$; see Figure 20). There was also a significant difference in temperature values before and after each testing session at each time-of-day ($F(1,23) = 58.96$, $p<0.001$). No interactions were observed (see Appendix 3(a) for details of this analysis).

From Figure 20 it can be seen that temperature gradually increased from its' lowest value at 0900 to its' highest value at 1800. The increase in temperatures after each test session can also be seen. To explore the time-of-day variations in temperature Newman-Keuls analyses were carried out to examine exactly where significant differences lay. These analyses showed that temperature values differ significantly from one time of the day to the next.

Figure 20: Mean Temperature Values Before and After Each Time of Day



Comparison of the diurnal trend seen for performance with that seen for temperature (see Figures 18 and 20) shows no evidence of a relationship between the two. Further, while the Newman-Keuls analyses showed consistent significant differences in temperature over the day, this was not mirrored in the Newman-Keuls analyses for performance.

To summarise, Experiment 3 showed response times were significantly faster at 1200 and 1800 than they were at 0900. Thus significant differences lay between early morning (0900) and later in the day (1200/1800). Response times were fastest when the icons were simple and icon concreteness did not significantly affect performance. Participants' temperatures significantly varied according to the time-of-day but no relationship between temperature and performance was apparent.

Experiment 4:Time-of-day effects in a semantic memory task

Experiment 4 used a search and match task using icons that varied in terms of their complexity and concreteness.

5.4 Method

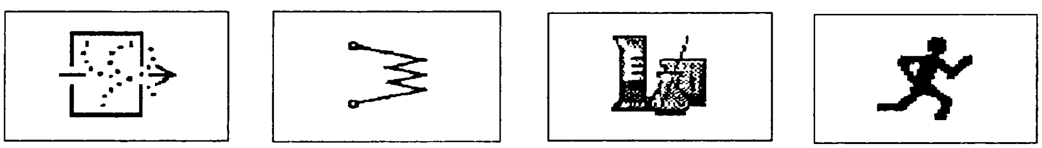
The paradigm used in Experiment 4 was identical to Experiment 3. This experiment differed from Experiment 3 in only one component of the task, participants were asked to match the function to the icon and this involved a semantic memory component (see Figures 17 and 22).

5.4.1. Participants

Twenty-four undergraduates and postgraduates from the University of Wales Swansea participated, four were male and twenty were female. The age range of participants in Experiment 6 was 18 to 25 years; the mean age was 20 years 9 months (standard deviation, 2 years 6 months). Some received course credit while others received a payment of £10.

5.4.2. Materials & Apparatus

The materials and apparatus used for Experiment 4 were identical to those used for Experiment 3, except that function labels were attached to each icon (see Figure 21).



Rinse *Spark Coil Ignition* *Chemistry* *Go Rapidly*

Figure 21: Examples of Each Icon Type and their Function Labels Used in Experiment 4

5.4.3. Design

The design of Experiment 4 was identical to that for Experiment 3.

5.4.4.Procedure

The procedure for Experiment 4 was identical to that for Experiment 3, except a target function label rather than an icon was presented on screen and the icon that represented the function label had to be found among an array of 9 icons that were subsequently shown on the screen (see Figures 22). As before, the target function label remained on screen for two seconds before the distractor array appeared, this seemed sufficient as research has found that adults read single words in 250-550 milliseconds (Waters, Seidenberg and Bruck 1984; Coltheart and Rastle 1994). Participants were given three opportunities to choose the correct icon as they were basically being asked to guess the matching icon during the early trials. If after the third attempt the correct icon still had not been chosen, all icons in the array apart from the correct one disappeared, the correct icon remained on screen for two seconds before the program automatically moved on to the next trial.

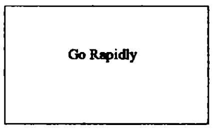


Figure 22 (a): Screen 1/Experiment 4, *target function label appears for two seconds before the distractor array appears.*

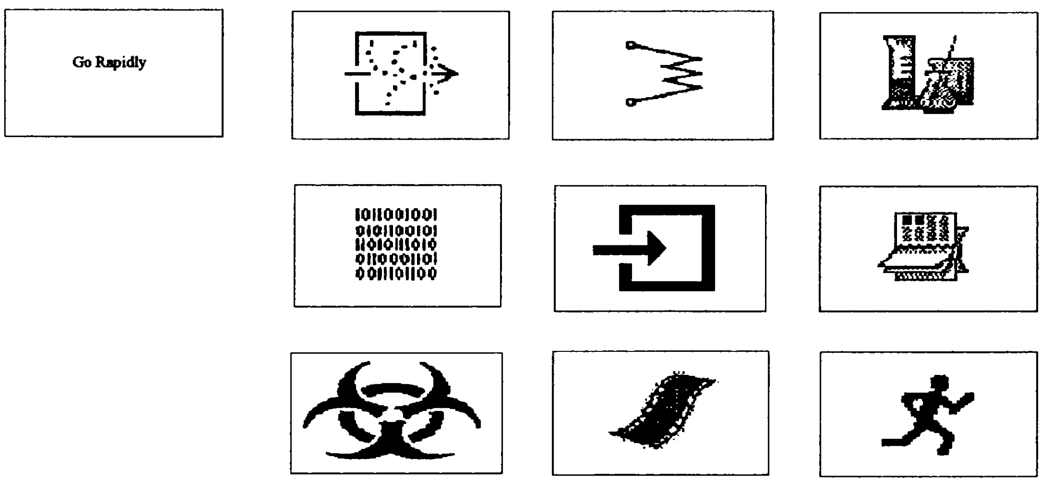


Figure 22 (b): Screen 2/Experiment 4, *target function label remains on screen as the distractor array appears. In this instance, the participant should use the mouse to click on the icon representing a person running.*

The presentation and randomisation of icons was identical to that for Experiment 3.

5.5 Results

5.5.1. Response Times

Table 7 shows that response times are reduced when icons are concrete. Response times were also faster for simple icons. It can also be seen that response times generally became faster across the 3 blocks of trials for each icon type at each time-of-day. From Figure 23 it can be seen that it is the diurnal trend seen for abstract-complex icon type that most closely resembles the diurnal trend seen in

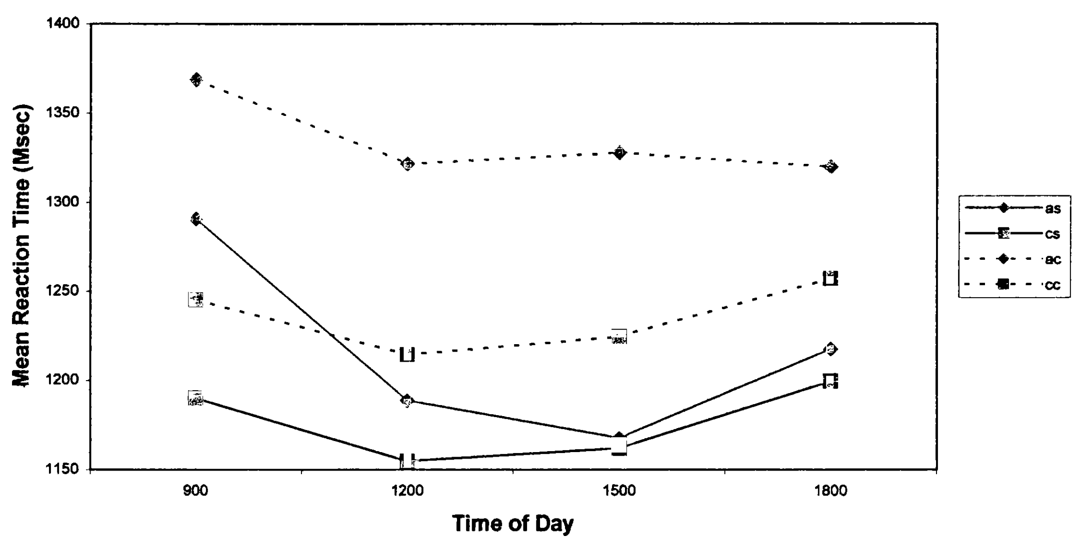
TOD	*Condition & Block Number*											
	***ASb1**	**ASb2**	**ASb3**	**ACb1**	**ACb2**	**ACb3**	**CSb1**	**CSb2**	**CSb3**	**CCb1**	**CCb2**	**CCb3**
0900	1360.0	1277.13	1236.62	1469.61	1294.77	1337.31	1203.27	1193.59	1173.84	1292.78	1231.46	1211.54
	(373.58)	(343.58)	(242.73)	(345.55)	(214.04)	(242.11)	(204.48)	(171.63)	(195.60)	(194.99)	(151.47)	(206.78)
1200	1222.77	1168.15	1176.08	1400.91	1274.67	1291.26	1167.50	1151.23	1146.04	1224.36	1199.35	1218.45
	(274.14)	(264.35)	(236.90)	(333.83)	(263.07)	(272.14)	(210.20)	(163.42)	(207.09)	(203.58)	(203.77)	(214.13)
1500	1184.79	1163.22	1155.93	1357.06	1318.69	1307.99	1193.87	1132.42	1160.24	1255.76	1238.45	1178.52
	(206.42)	(195.47)	(153.55)	(209.64)	(215.58)	(242.54)	(151.89)	(173.99)	(173.37)	(139.86)	(185.49)	(170.54)
1800	1283.86	1177.79	1201.62	1408.28	1288.58	1270.02	1213.51	1166.82	1173.77	1291.49	1251.95	1227.17
	(223.66)	(211.57)	(206.28)	(259.50)	(199.42)	(208.41)	(312.83)	(128.74)	(184.74)	(169.33)	(177.65)	(193.63)

Table 7: Mean (& standard deviations) of Response Times for Each Condition at Each Time-of-Day

*AS = abstract-simple; AC = abstract-complex; CS = concrete-simple; CC = concrete-complex previous experiments, with all other icon types showing a different pattern to those previously seen.

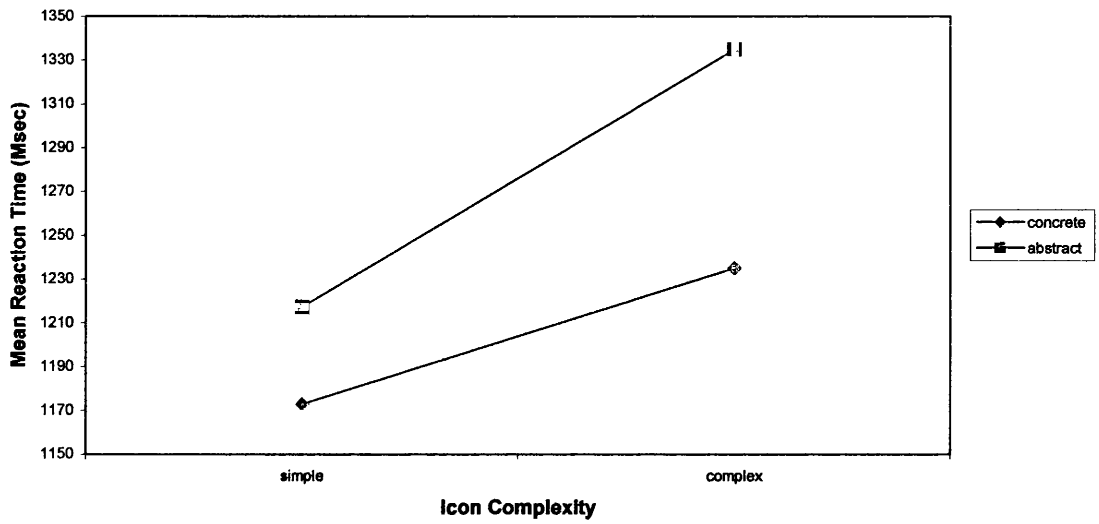
A four-factor repeated measures analysis of variance was carried out to examine the effects of the time of day (0900, 1200, 1500, 1800), icon concreteness (abstract versus concrete), icon complexity (simple versus complex) and block number (1, 2, 3). See Appendix 3 for a full summary of results. There was no variation in performance in accordance with time-of-day ($F(3,69) = 1.14$, p = ns; see Figure 23). However, response times were significantly faster for concrete, rather than abstract, icons ($F(1,23) = 14.14$, $p < 0.001$) and for simple, rather than complex, icons ($F(1,23) = 53.89$, $p<0.001$). Also, performance significantly varied over blocks of trials ($F(2,46) = 24.80$, $p<0.001$), with response times becoming faster across the 3 blocks of trials at each time-of-day (see Appendix 3(a) for details of this analysis).

Figure 23: Mean Reaction Times for Each Icon Type at Each Time of Day



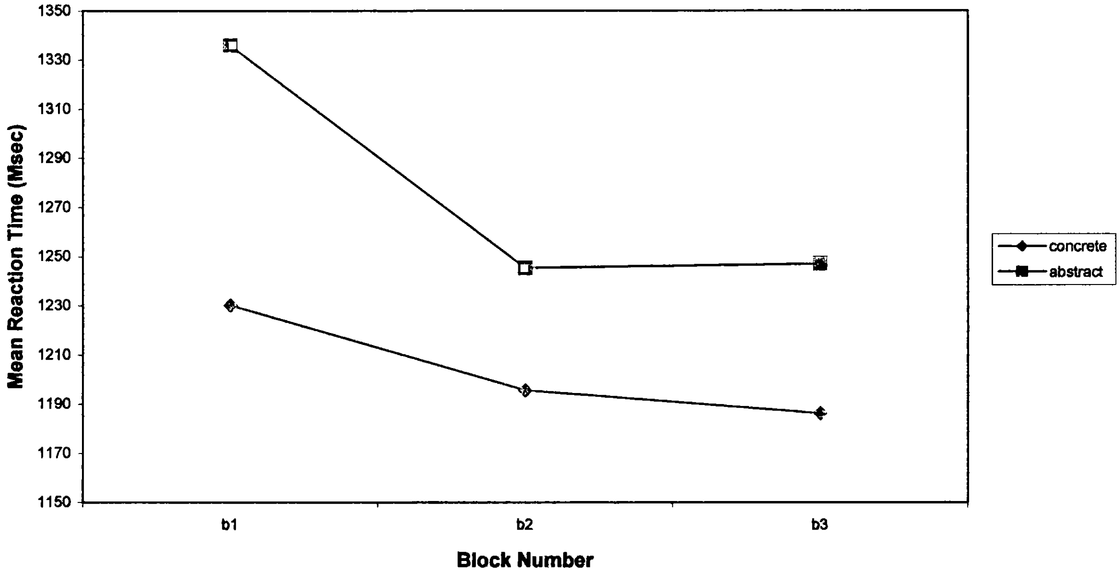
A significant interaction was observed between icon concreteness and icon complexity ($F(1,23) = 6.45$, $p<0.05$). Figure 24 shows response times were faster for both concrete and abstract icons when the icons were also simple. From Figure 24 it can be seen that the difference in response times between concrete and abstract icons was greater when the icons were complex. Accordingly, simple main effects revealed the difference in response times for concrete and abstract simple icons was not significant ($F(1,23) = 4.07$, $p = 0.056$), while the difference in response times for concrete and abstract complex icons was significant ($F(1,23) = 20.31$, $p<0.001$; see Appendix 3(b) for details of this analysis).

Figure 24: Interaction Between Icon Concreteness and Icon Complexity



A further significant interaction was observed between icon concreteness and block of trials ($F(2,46) = 4.48$, $p<0.05$). Figure 25 shows that as participants progressed through the blocks of trials the response times decreased, this decrease in response times seemed to begin more quickly for abstract icons. Simple main effects revealed significant reductions in response times for concrete ($F(2,46) = 5.67$, $p<0.01$) and abstract ($F(2,46) = 23.38$, $p<0.001$) icons as participants progressed from block 1 to block 3. This difference in reaction times between concrete and abstract icons was significant at block 1 ($F(1,23) = 14.48$, $p\leq 0.001$), at block 2 ($F(1,23) = 7.19$, $p<0.05$) and block 3 ($F(1,23) = 9.80$, $p< 0.001$), however the strength of this effect became less as blocks progressed (see Appendix 3(b) for details of this analysis). No other interactions were significant.

Figure 25: Interaction Between Icon Concreteness and Block Number



5.5.2. Accuracy

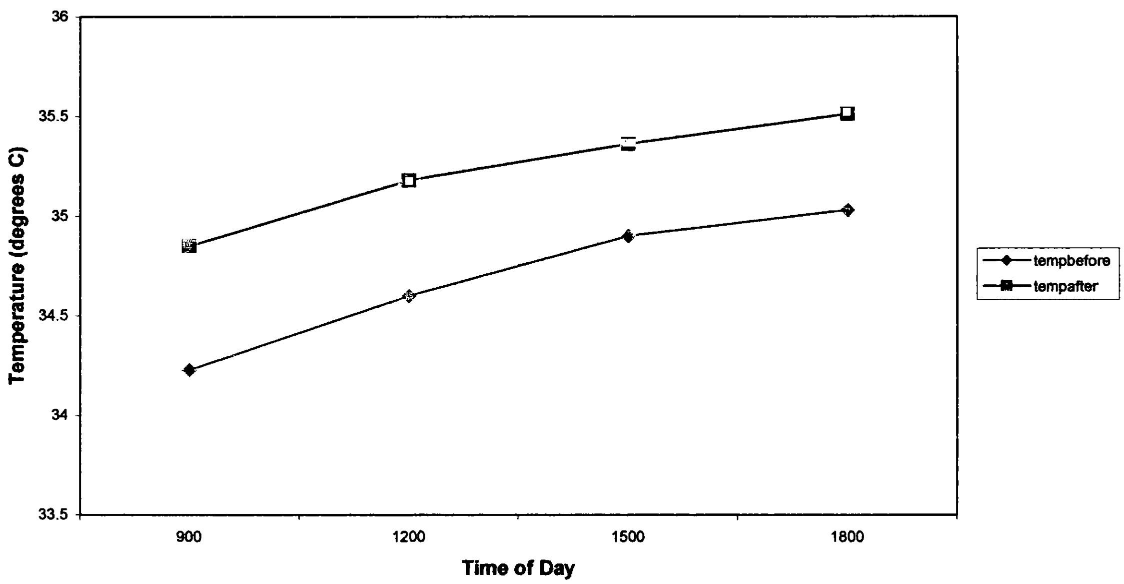
Slightly more variation in the accuracy data obtained was found for Experiment 4 than in Experiment 3. For full details of percentage accuracy rates see Appendix 3(c). Mean percentage accuracies for each participant ranged from 92.59% to 100% with the mode being 99.31%. Mean percentage accuracies for each time-of-day ranged from 97.03% at 1200 to 99.70% at 0900. The overall mean percentage accuracy was 97.37%. Since the mean accuracy levels were universally high, no further analyses were conducted.

5.5.3. Temperature

A two-factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900, 1200, 1500, 1800). Significant differences in temperature values were found as a result of the time-of-day ($F(3,69) = 42.82$), $p<0.001$; see Figure 26). There was also a significant difference in temperature values before and after each testing session at each time-of-day ($F(1,23) = 74.85$, $p<0.001$). No interactions were observed (see Appendix 3(a) for details of this analysis).

From Figure 26 it can be seen that as in previous experiments, temperature gradually increased over the course of the waking day. The tendency for temperatures to increase after each test session can also be seen. To explore the time-of-day variation in temperature further, Newman-Keuls analyses were carried out to examine exactly where significant differences lay. These analyses showed that as before, temperature values differed significantly from one time of the day to the next.

Figure 26: Mean Temperature Values Before and After Each Time of Day



Comparison of the diurnal trend seen for performance with that seen for temperature (see Figures 23 and 26) again shows no obvious relationship between the two.

5.6 Discussion

Experiment 3 (summarised earlier) showed that response times varied significantly according to the time-of-day. In contrast response times in Experiment 4 did *not* vary significantly as a result of the time-of-day. However, abstract-complex icons showed a trend that closely resembled those seen previously, suggesting it is this icon type that is most important in time-of-day performance trends. Responses were fastest when the icons were simple and when they were concrete. Performance was found to improve over the three blocks of trials completed at each time point and response times decreased over the three blocks more rapidly for abstract icons. As in previous experiments, participants' temperatures significantly varied according to the time-of-day with temperature increasing from early morning to early evening, but no evidence of a relationship between temperature and performance was apparent.

5.6.1. Time-of-Day Trends and Previous Research

The time-of-day trends observed in Experiments 3 and 4 differed considerably from those observed in Experiments 1 and 2. Experiment 3 attempted to establish if time-of-day effects can be observed in a simple search task, changes in response times as a result of the time-of-day were observed and responses were slowest at 0900. Although this was consistent with Experiments 1 and 2 in which simpler tasks, with a lower memory load, resulted in slightly poorer performance early in the day when arousal was low, it is important to note that the performance decrements seen at this time in Experiment 3 were far greater than those decrements seen under the lower memory load conditions in Experiments 1 and 2. As Experiment 3 involved little or no memory load support is provided for the notion that memory load is the key to improved early morning performance. The lack of a memory component here, resulted in marked performance decrements at 0900 that were significantly slower than performance at 1200 and 1800.

In Experiment 3, 0900 was consistently slower than 1500. Moreover, although 1500 showed a decline in performance from 1200, this difference was not significant and no significant improvement in performance after 1500 was observed. Therefore no support was found for the post-lunch dip (Owens et al, 2000; Lenne et al, 1997). This suggests that the post-lunch dip can be manipulated by task demands, a finding that is consistent with previous research that has shown the post-lunch dip to be flexible. Such research has found for example that noise removes the fall in performance while a high carbohydrate meal exacerbates it (Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b; Smith et al, 1990; Smith, 1988). Thus, it would appear that exact demands of the task might be of great importance in determining the exact diurnal trend seen (Smith, 1992).

Experiment 3 also aimed to determine whether the observed diurnal trend varied according to icon type. This was not found to be the case with no interactions being observed between time-of-day and icon type and with all icon types showing a similar performance trend over the day. However, examination of Figure 18 shows time-of-day effects may be more pronounced for abstract icons. This suggests that it is the more difficult icons, that are not closely mapped to the object or function that they are intended to represent (Rogers, 1989), that may be more susceptible to diurnal performance fluctuations. Nonetheless, while it is interesting to note the above, the evidence is minimal, thus no conclusions can be drawn from this at present.

The findings of Experiment 4 differed markedly from Experiment 3. In Experiment 4, which attempted to establish if a time-of-day effect exists in a task where meaning must be attached to an icon, no consistent, nor significant, time-of-day trends emerged (see Figure 23). Since the major difference between Experiments 3 and 4 was the fact that participants had to learn to match the icon to its' function in Experiment 4 (see Figure 22), while participants were simply required to search for the matching icon in Experiment 3 (see Figure 17), it seems likely that the different time-of-day trends observed are the result of the semantic memory component in Experiment 4.

Since Experiment 4 involved semantic memory one might expect the results to be consistent with previous semantic memory literature that has found performance on these tasks to improve later in the day (for example, Millar et al, 1980, Tilley and Warren, 1983, Smith, 1987a). The results of Experiment 4 show no consistent evidence for this. Further, previous research has found semantic access to be slowest mid-afternoon (Whitney and Williams, unpublished), yet Experiment 4 does not support this. However, research has also shown that changing the nature of semantic memory tasks can alter the observed time-of-day trend. For example, Smith (1987a) changed the nature of his semantic memory task and a previously observed time-of-day effect disappeared. Consequently it may be a specific aspect of the Experiment 4 task resulting in the absence of a time-of-day trend. However, at present, the precise reasons for the lack of a time-of-day effect in Experiment 4 are unclear.

Diurnal Trends in Performance and the Arousal Model and Other Mechanisms

The difference in the timing of peak performance between difficult (Experiments 1 and 2) and easier (Experiment 3) tasks can be explained in terms of arousal levels and it appears that it is the memory load involved in a task (determined by the visual memory requirement, difficulty of response and difficulty of icon discrimination) that determines when the peak in performance will occur. Thus Experiment 3 has provided further support for the arousal theory framework, which proposed that the high arousal levels are best for low memory load tasks. In contrast however, performance again failed to parallel the temperature trend and this does not support this framework. Also this does not support Monk's (1982) theory that low memory load tasks are mediated by an arousal rhythm that parallels the temperature trend. This suggests that theories proposing that time-of-day trends are the result of input from homeostatic process and from a circadian timing system (for example, Monk et al, 1983; Folkard and Akerstedt, 1992) may be more accurate.

The memory load involved in the task used in Experiment 4 was difficult to ascertain, although it was a semantic memory task and so participants must learn, and thus remember, the meaning of the icons the function labels remain on screen during the search thus participants did not need to remember which function label had been shown. Further, the difficulty participants had in learning/remembering the meaning of the icons here may be related to intelligence and prior experience, neither of which were examined. Consequently, as it was difficult to determine the memory load involved in Experiment 4 it was difficult to relate the findings to the arousal theory framework. However, once again performance did not significantly correlate with temperature, failing to support the assertion of this framework that performance is related to temperature. However, it was plausible that due to the nature of the task in Experiment 4, participants were changing their strategies between early morning and early evening as their level of arousal changed. Research has found more 'maintenance processing' to be used in the morning, where attention focuses more on the physical characteristics while more 'elaborative processing' is used in the afternoon when attention focuses more on the meaning of the stimuli (Folkard, 1979, 1980, 1983; Lorenzetti and Natale, 1996; Oakhill, 1986a, 1986b, 1988; Marks and Folkard, 1988). This explanation may account for the poor performance observed at 0900 relative to the rest of the day in this task, however performance generally declined again at 1800, which is not consistent with this explanation.

Effects of Icon Concreteness and Complexity

Both experiments used icon sets in which concreteness and complexity were varied orthogonally (see Figure 2). In both experiments the effects of these icon characteristics were apparent. Response times were significantly faster for simple icons. This is in accordance both with Experiments 1 and 2 and with previous research (Byrne, 1993; McDougall et al, 2001). As noted earlier this is largely because simple icons reduce visual search time (McDougall et al, 1996) while search time increases for complex icons. Interestingly, response times were significantly faster for concrete icons in Experiment 4, while icon concreteness did not affect performance in Experiment 3. This is the result of different task demands. Since concreteness and meaning are closely related (Paivio et al, 1968; Gilhooly and Logie, 1980), Experiment 3, a basic visual search task that did not require any associations with meaning, showed no effect of concreteness. The effects of concreteness became apparent only in Experiment 4, where meaning was important. Similar effects have been observed by McDougall et al (2000) who found that icon concreteness had no effect on response times in a search task such as that used in Experiment 3, but were important in determining performance in a search and match task such as that used in Experiment 4.

It was noted in the introduction that the icon set used in Experiments 3 and 4 may be more distinctive than those used in Experiments 1 and 2. However, Boersema and Zwaga (1996) found practice is required for the 'pop-out' of more distinctive stimuli to occur, whether participants' would have had enough practice on the task for this effect to occur is not known, thus all that can be concluded is that it is possible that icon distinctiveness had an effect.

In both experiments participants completed three blocks of trials at each time-of-day, in Experiment 4 response times became significantly faster over the three blocks while in Experiment 3 this trend did not quite reach significance. This is attributable to practice effects and it is likely that it did not reach significance in Experiment 3 due to the easiness of the task, while the more difficult task in Experiment 4 showed more rapid improvements as experience increased. Moreover, in Experiment 4 abstract icons showed a more rapid improvement over blocks than concrete icons and the performance gap between abstract and concrete icons had closed markedly by block 2, this demonstrates the short-lived advantage of concrete icons described by McDougall et al (2001) where the meaning of icons becomes less important as the icon set is learned.

5.6.2. Experiment 5

Attention now focused on the tasks used in the first three experiments and subsequent experiments attempted to determine what aspects of these tasks and/or features of icons were important in determining the exact diurnal trend observed. There are three main differences between the tasks used in Experiments 1 and 2 and that used in Experiment 3 that each contribute to the memory load involved in the task. Experiments 1 and 2 required participants to remember the target icon, while in Experiment 3 the icon remained on screen and did not need to be remembered, thus the visual memory component was absent from Experiment 3. Also, Experiments 1 and 2 required participants to state whether an icon was present in the distractor array or not by making a 'yes' or 'no' keyboard response, this necessarily meant that the icon *was not* always present in the distractor array. In Experiment 3 participants were simply required to click the matching icon with the mouse and the target *was* always present in the display. Thus response type also differed between Experiment 3 and Experiments 1 and 2. Furthermore, the icons in Experiment 3 appeared to be more distinctive than those in Experiments 1 and 2. Thus it is not clear what the difference in diurnal trends between Experiments 1 and 2 and Experiment 3 can be attributed to. The following experiments explored the role of each task component. Thus, Experiment 5 was identical to Experiment 3 with the exception that the visual memory component found in Experiments 1 and 2 was reintroduced.

5.6.3. Conclusions

It can be concluded that individuals who are required to deal with icons, will be prone to similar diurnal performance fluctuations regardless of icon type, although abstract icons may result in more pronounced time-of-day effects. Further, if the task involves little or no working memory component marked performance decrements can be expected at 0900. If the task involves semantic memory, individuals can perhaps expect to be free from diurnal performance fluctuations. However, the exact demands of a semantic memory task need to be carefully controlled to avoid the emergence of diurnal performance fluctuations.

Chapter 6

Experiment 5: The Effect of Visual Memory

6.1 Introduction

Experiment 3 provided evidence that diurnal trends in performance occur in a simple icon task where a simple basic search process is involved. This experiment also provided evidence that all icon types show similar performance trends over the day. However, the exact nature of the performance trend seen differed from that seen in Experiments 1 and 2. There were three differences between the experiments, that each contributed to the task difficulty/memory load involved, that are possible reasons for this:

(i) *Visual Memory Component*

Experiments 1 and 2 involved a visual memory component. This was because the target icons were not displayed during search of the array and had to be remembered for search to be successful. In contrast, Experiment 3 required participants to match an icon in the array while the target icon was still present.

(ii) *Difficulty of Response*

Also a 'yes'/ 'no' keyboard response was used in Experiments 1 and 2 where participants made a 'yes' keyboard response if the target was present among the distractors or a 'no' keyboard response if it was not. In contrast, Experiment 3 simply required participants to click on the icon that matched the target using the mouse.

(iii) *Difficulty of Icon Discrimination*

Experiments 1 and 2 used icons that were specifically designed for those experiments (see Figures 6 and 11), these icons were non-distinctive and therefore difficult to discriminate between. Meanwhile, Experiment 3 used icons that had been orthogonally varied in terms of their concreteness and complexity (see Figure 16). This icon set was perceptually different being more distinctive and therefore easier to discriminate.

Experiment 5 examined the possibility that the visual memory requirement in Experiments 1 and 2 and McFadden et al's (1997) study, which was absent from

Experiment 3, produced the change in the pattern of time-of-day effects observed in Experiment 3. As Rogers (1989) highlighted, the effect of visual memory in icon tasks has been largely ignored, perhaps because human-interface interaction does not usually require the user to remember the icons. However, research suggests this could be an important factor in icon tasks. For instance, iconic memory has been found to have a limited capacity (Coltheart, 1980). Furthermore, Rogers (1988) found memory for the meaning of *icons* improved over time while it remained the same for *labels,* participants were found to have more difficulty in remembering what the labels meant but had little difficulty remembering what the icons meant. Rogers (1989) explained these findings in terms of Paivio's (1971, 1986) dual coding theory suggesting that the meaning of icons are likely to be better remembered because the pictorial information is stored as imagens and logogens, that is they are stored both in visual and verbal memory stores. This aids recognition and/or recall as information can be accessed from either one of the two stores, or both. The icons employed in Experiment 3, and the use of a mouse click response type, remained, thus varying only the visual memory component in the task involved in Experiment 5.

6.1.1. Aims

Experiment 5 examined the following:

1) Whether the diurnal performance trend seen for Experiment 3 was altered by changing task demands through the reintroduction of a visual memory component.
2) How the diurnal trend produced compared with the trend found for Experiment 3. If visual memory is the key then we might expect a pattern of performance more similar to that for Experiments 1 and 2, conversely if visual memory is not the key then we might expect a pattern of performance more similar to that for Experiment 3.
3) Experiment 5 examined if the temperature trend showed a relationship to the performance trend.
4) As before Experiment 5 also determined whether differences occurred in the speed/accuracy of response for complex versus simple and abstract verses concrete icons in a way similar to that previously seen in Experiments 1 – 4.

As Experiments 1, 2 and 3 showed significant diurnal performance trends, and Experiment 5 used a similar paradigm, it was expected that a significant time-of-day trend would be observed in Experiment 5. Further, as time-of-day effects have been shown to be sensitive to exact task demands (Smith, 1992), it was expected that, as Experiment 5 involved a visual memory component not used in Experiment 3, the diurnal trend produced in this experiment would not be identical to that seen in Experiment 3, but would be similar to that seen for Experiments 1 and 2, which also involved a visual memory component. Additionally, the post-lunch dip in performance (for example, Owens et al, 2000; Lenne et al, 1997) at 1500 was expected to reappear as a result of the task demands more closely resembling those used in Experiment 2, which showed a post-lunch dip in performance. Alternatively however, it was possible that the time-of-day trend produced for Experiment 5 would differ from that seen for all previous experiments or the trend may be identical to that seen for Experiment 3, in this case it is possible that the 'yes'/ 'no' response type used in Experiments 1 and 2, but omitted from Experiment 5, is the intervening factor.

As in previous experiments, it was expected that the speed and accuracy of response would be superior for simple icons (McDougall, et al 1996; Byrne, 1993), while icon concreteness was not expected to significantly influence performance.

6.2 Method

The paradigm used in Experiment 5 was identical to Experiment 3, with the exception that participants were required to remember the target icon that disappeared from screen after two seconds before the distractor array, from which participants had to choose the matching icon, appeared.

6.2.1 Participants

Twenty-four undergraduates and postgraduates from the University of Wales Swansea participated, eight were male and sixteen were female. The age range of participants was 18 to 24 years; the mean age was 20 years 2 months (standard deviation, 1 year 7 months). Some received course credit for their participation while others received a payment of £10.

6.2.2 Materials & Apparatus

The materials and apparatus used for Experiment 5 were identical to those used for Experiment 3.

6.2.3 Design

Twenty-four participants were used in total for Experiment 5. The participants were divided into four groups of four and into four groups of two. Four participants were tested at a time. Each participant was tested once at each of the following times: 0900; 1200; 1500; 1800. Each participant also completed a practice session at one of these times. The practice session a participant attended, and the order of testing thereafter, was counterbalanced using a Cyclic Latin Square (see Table 8) to ensure the administration of the experimental conditions was balanced across participants. Due to time and equipment constraints, two counterbalanced cycles were used; participants were firstly divided into 4 groups of 4 for one cycle and into 4 groups of 2 for the second cycle. One group was tested at a time during the first cycle, while two groups were tested at a time during the second cycle. In this way 6 participants completed each different order of testing in the minimum amount of time. Each participant's sessions were completed consecutively in approximately 24 hours.

Group	*Practice*	*Test 1*	*Test 2*	*Test 3*	*Test 4*
1	0900	1200	1500	1800	0900
2	1200	1500	1800	0900	1200
3	1500	1800	0900	1200	1500
4	1800	0900	1200	1500	1800

Table 8: Cyclic Latin Square

6.2.4 Procedure

To start each experimental trial, participants were required to click on an "ok" button present on the screen using a mouse. Once participants had clicked the "ok" button it turned grey and the target icon appeared on screen for a period of two seconds before disappearing. Immediately, the array of 9 distractor icons appeared on the screen from which participants were required to select the target icon as quickly as possible using the mouse. Timing began when this distractor array appeared. Participants were instructed to respond as quickly and as accurately as they could. Timing finished as soon as one of the icons in the display was selected. The next trial would then begin and the whole procedure was repeated. No feedback was given regarding the correctness of the response. If no response was given within 6 seconds, the program moved on to the next trial automatically (see Figure 27). This was repeated for each of the three blocks. There was a two-minute break between each block.

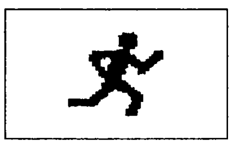


Figure 27 (a): Screen 1/Experiment 5, *target icon appears on screen for two seconds before disappearing.*

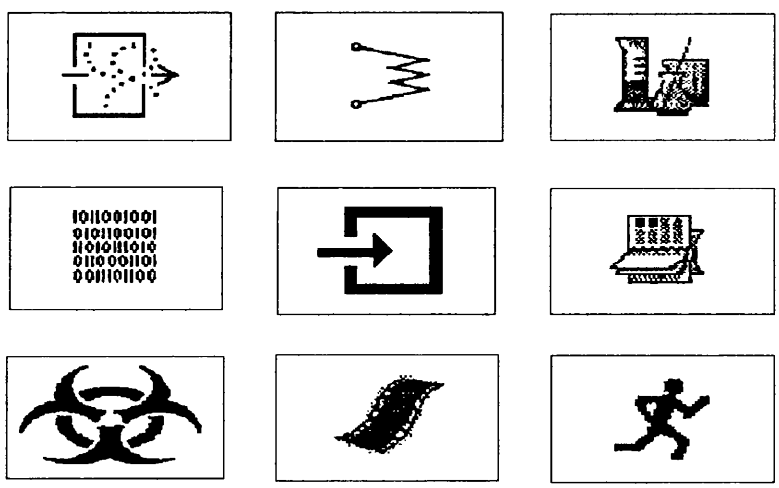


Figure 27 (b): Screen 2/Experiment 5, *immediately the distractor array appears, the target icon must be memorised and identified among the distractors, the participant should click on the matching icon using the mouse.*

Eardrum temperature readings were taken before and after each test session and were taken from the right ear on every occasion. The procedure was identical at each test session.

The presentation and randomisation of icons was identical to that for Experiment 3.

6.3 Results

6.3.1. Response Times

Table 9 illustrates a trend in performance that varied according to the time-of-day where generally reaction times were fastest at 1200 or 1500 and slowest at 0900. Performance generally improved between 0900 and 1200 before falling again at 1500. It is also shown that generally reaction times were fastest when simple icons were presented. Finally, it can be observed from Table 9 that abstract icons generally produced faster reaction times than concrete ones when the icons were also simple and conversely, concrete icons were faster than abstract ones when the icons were also complex.

A four-factor repeated measures analysis of variance was carried out to examine the effects of the time-of-day (0900, 1200, 1500, 1800), icon concreteness (abstract versus concrete), icon complexity (simple versus complex) and block number (1, 2, 3). See Appendix 4 for a full summary of results. The effect of the time-of-day was significant ($F(3,69) = 5.17$, $p<0.01$; see Figure 28). Responses were fastest when icons were simple and slowest when icons were complex ($F(1,23) = 95.68$, $p<0.001$). Icon concreteness did not significantly affect performance ($F(1,23) = 0.97$, p = ns) nor did block number ($F(2,46) = 0.17$, p = ns) (see Appendix 4(a) for details of this analysis).

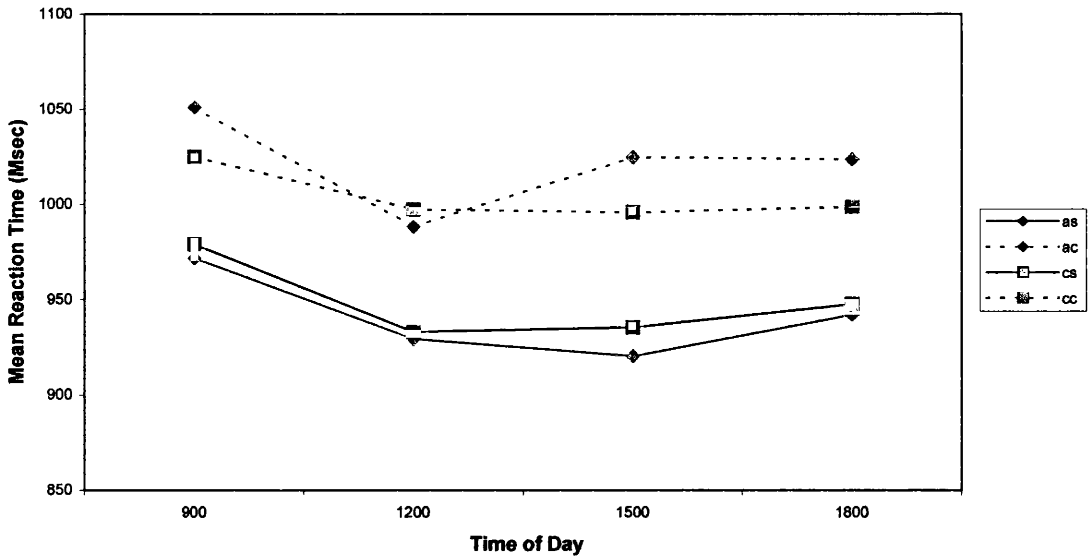
From Figure 28 it can be seen that the exact diurnal trend observed varied according to icon type. Once again it was the abstract-complex icon type that revealed a trend most similar to that seen previously, where response times were fastest at 1200 and slowest at 0900. The remaining icon types however showed similar trends to each other where responses were slowest at 0900 while the remaining times of the day showed minimal difference from each other. In order to explore the time-of-day trends seen for each icon type further, Newman-Keuls analyses were carried out to examine exactly where significant differences lay. Overall, these analyses revealed significant differences in response times between 0900 and 1200, 0900 and 1500 and between 0900 and 1800. This suggests that time-of-day effects are most marked between early morning (0900) and the rest of the day (1200/1500/1800).

TOD	*Icon Type & Block Number*											
	***ASb1**	**ASb2**	**ASb3**	**ACb1**	**ACb2**	**ACb3**	**CSb1**	**CSb2**	**CSb3**	**CCb1**	**CCb2**	**CCb3**
0900	965.46 (136.75)	989.21 (142.02)	960.83 (168.34)	1032.88 (158.72)	1064.13 (163.86)	1055.79 (151.13)	988.08 (157.14)	963.75 (136.69)	985.38 (162.39)	996.13 (140.33)	1044.71 (153.04)	1034.25 (166.51)
1200	929.33 (182.49)	948.83 (131.06)	910.33 (141.57)	1006.83 (196.12)	985.79 (116.75)	972.46 (140.32)	929.33 (146.88)	945.54 (138.37)	924.79 (127.01)	1006.21 (157.90)	1001.29 (111.85)	984.63 (157.92)
1500	930.04 (138.63)	912.38 (97.90)	919.04 (102.74)	1020.54 (154.19)	1021.83 (147.33)	1031.54 (197.30)	945.29 (132.07)	914.83 (134.07)	946.38 (153.31)	973.54 (151.45)	1008.08 (133.82)	1005.79 (144.13)
1800	951.33 (163.62)	944.17 (145.23)	930.71 (145.85)	1015.25 (184.18)	1025.38 (172.70)	1030.13 (173.81)	963.67 (163.52)	934.38 (130.93)	944.92 (136.04)	1013.58 (146.96)	989.75 (149.32)	992.29 (151.26)

Table 9: Mean (& standard deviations) Response Times for Each Icon Type at Each Time of Day

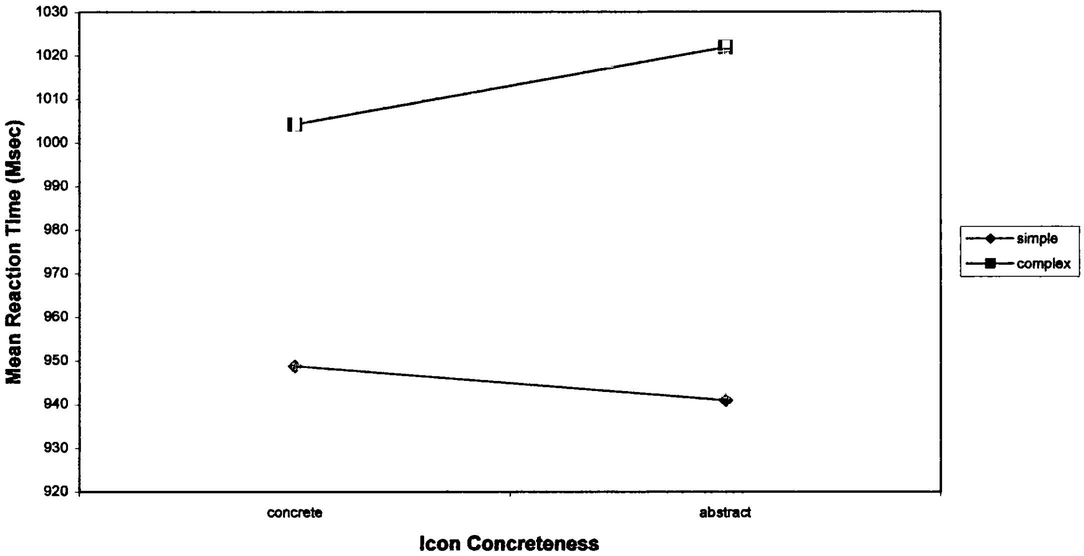
*AS = abstract-simple; AC = abstract-complex; CS = complex-simple; CC = concrete-complex

Figure 28: Mean Reaction Times for Each Icon Type at Each Time of Day



A significant interaction was observed between icon concreteness and icon complexity ($F(1,23) = 5.20$, $p<0.05$) on response times. Figure 29 shows response times decreased when simple abstract icons were shown but increased when abstract complex icons were shown. From Figure 29 it can be seen that the difference in response times between simple and complex icons became greater when the icons were abstract. Nevertheless, simple main effects revealed that the difference between simple and complex icons was significant for both abstract ($F(1,23) = 51.98$, $p<0.001$) and concrete icons ($F(1,23) = 90.09$, $p<0.001$; see Appendix 4(b) for details of this analysis). No other interactions were significant.

Figure 29: Interaction Between Icon Concreteness and Icon Complexity



6.3.2. Accuracy

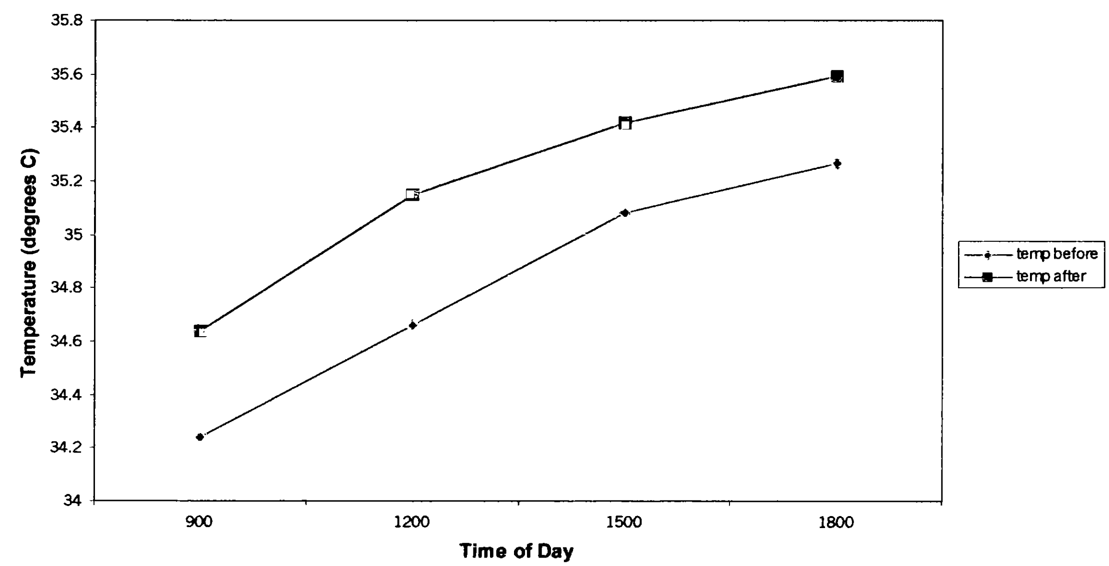
Due to the nature of the task involved in Experiment 5 there was very little variation in the accuracy data obtained. For full details of percentage accuracy rates see Appendix 4(c). Mean percentage accuracies for each participant ranged from 91.20% to 100% with the mode being 100%. Mean percentage accuracies for each time-of-day ranged from 99.00% at 1500 to 99.36% at 1800. Overall the mean percentage accuracy over all times of day and over all participants and icon types was 99.24%. Due to the presence of these ceiling effects no further analyses were conducted.

6.3.3. Temperature

A two-factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900, 1200, 1500, 1800). Significant differences in temperature values were found as a result of differences in the time-of-day ($F(3,69) = 69.69$, $p<0.001$; see Figure 30). There was also a significant difference in temperature values before and after each testing session at each time-of-day ($F(1,23) = 76.32$, $p<0.001$) with temperature being higher after each testing session than they were before each session. No interactions were observed (see Appendix 4(a) for details of this analysis).

From Figure 30 it can be seen that again, temperature gradually increased over the course of the waking day with the lowest temperature values being seen at about 0900 and the highest at about 1800. Figure 30 also illustrates the increase in temperature at the end of each test session. To explore the time-of-day variations in temperature, Newman-Keuls analyses were carried out to explore exactly where significant differences lay. These analyses revealed that, as in previous experiments, all temperature values differed significantly from one time of the day to the next.

Figure 30: Mean Temperature Values Before and After Each Time of Day



Comparison of the time-of-day trend seen for performance with that seen for temperature (see Figures 28 and 30) shows no obvious parallelism between the two. Furthermore, Newman-Keuls analyses showed temperature to significantly differ between each time-of-day, yet this was not mirrored in the performance analyses.

6.4. Discussion

To summarise, response time varied in accordance with the time-of-day. Responses were slowest at 0900 for all icon types. It was found that responses were significantly different between 0900 and the rest of the day, all other times of the day showed very similar response times. Not surprisingly, responses were faster for simple icons. Icon concreteness did not significantly affect performance. Participants' temperatures significantly varied according to the time-of-day, increasing from a minimum at 0900 to a maximum at 1800, but no relationship between temperature and performance was apparent.

6.4.1. Time-of-Day Trends and Previous Research

Experiment 5 aimed to determine the effect of visual memory. The observed time-of-day trend was similar to that seen in Experiment 3. The abstract-complex icons revealed the most similar trend to that seen in previous experiments.

Response times in both, Experiments 3 and 5, were significantly slower at 0900 (see Figures 18 and 28). This suggests that even the introduction of a visual memory component keeps the memory load involved in the task relatively low, resulting in poor performance levels at 0900. This suggests that the memory load of the task is not high enough to result in improved performance seen at 0900 in Experiments 1 (see Figure 8) and 2 (see Figure 13). One of the best performance times in both experiments was 1200 and this is consistent with working memory literature (for example, Laird, 1925, Owens et al, 2000, Folkard, 1975). In accordance with Experiment 3, Experiment 5 does not support the notion of a post-lunch dip in performance (Owens et al, 2000; Lenne et al, 1997), again supporting the notion that the post-lunch dip is flexible (Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b; Smith et al, 1990; Smith, 1988). Finally, as in Experiment 3, examination of Figure 28 suggests that time-of-day effects may be more marked for abstract icons.

Thus, here the significant difference in responses again lay between 0900 and the rest of the day, essentially then, Experiment 5 is revealing the same trend as seen in Experiment 3. This suggests that although the introduction of a visual memory component may slightly increase memory load, its' effect is not of critical importance to the general time-of-day trend observed. This may be attributable to participants having little difficulty in remembering the icons (Rogers, 1988).

In sum, so far it seems that tasks with a high memory load (see Experiments 1 and 2) show improved performance at 0900 while tasks with a low memory load (see Experiments 3 and 5) show severe performance decrements at this time. This supports the notion that memory load is the key to superior early morning performance. Furthermore, this also suggests that memory load needs to be quite high in order for this superiority to occur, since it would appear from Experiment 5, that reintroducing a visual memory component alone does not increase memory load to the extent that performance early in the morning then improves. Consequently, it appears that it is response type or icon distinctiveness that is important.

It is interesting to note that the time-of-day trend seen for the concrete-complex, concrete-simple and abstract-simple icon types in Experiment 5 were beginning to show evidence of a reverse trend as seen in Experiment 4 (see Figure 31), where meaning had to be attached to an icon in a rather cognitively demanding task. Why this was the case is unclear. However, the abstract-complex icons have maintained the same general trend throughout. The answer probably lies in a statement made by Folkard and Hill (2002) that working memory tasks involve several different cognitive subsystems such as short-term storage of information, information processing and throughput, as such it is plausible that the observed time-of-day trends are an "outcome of a combination of different trends associated with the different cognitive mechanisms involved" (p.57). It is likely that the different task demands are utilising different cognitive subsystems to produce varying performance trends over the day.

So it seems that visual memory per se is not a vital demand of the task used here. However, aside from differences in icon distinctiveness and therefore difficulty of icon discrimination, there still remains a fundamental difference in the cognitive demands involved in each experiment that could account for the different diurnal trends seen between Experiments 1 and 2 and Experiment 3: difficulty of response. Clearly this aspect of the task demands requires further investigation.

Diurnal Trends in Performance and the Arousal Model and Other Mechanisms

Once again, support was provided for the notion from the arousal theory that the low arousal levels associated with the early morning result in poor performance on low memory load tasks. However, performance on this relatively simple task with a low memory load, again showed no relationship with the temperature trend and this is inconsistent with the arousal framework. This supports Carrier and Monks' (2000) suggestion that more contemporary theories (for example, Monk et al, 1983; Monk et al, 1989; Dijk et al, 1992; Johnson et al, 1992; Folkard and Akerstedt, 1992) may be more accurate.

Figure 31(a): Mean Reaction Times for Each Icon Type at Each Time of Day in Experiment 4

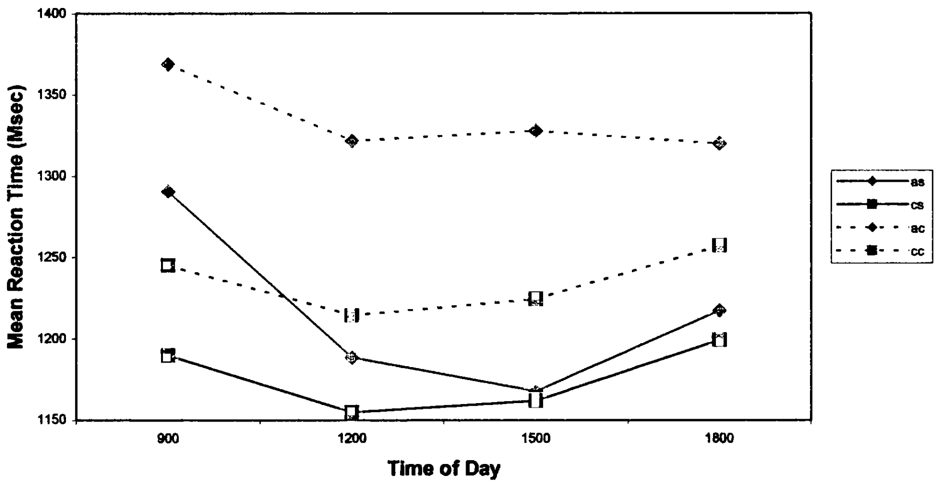
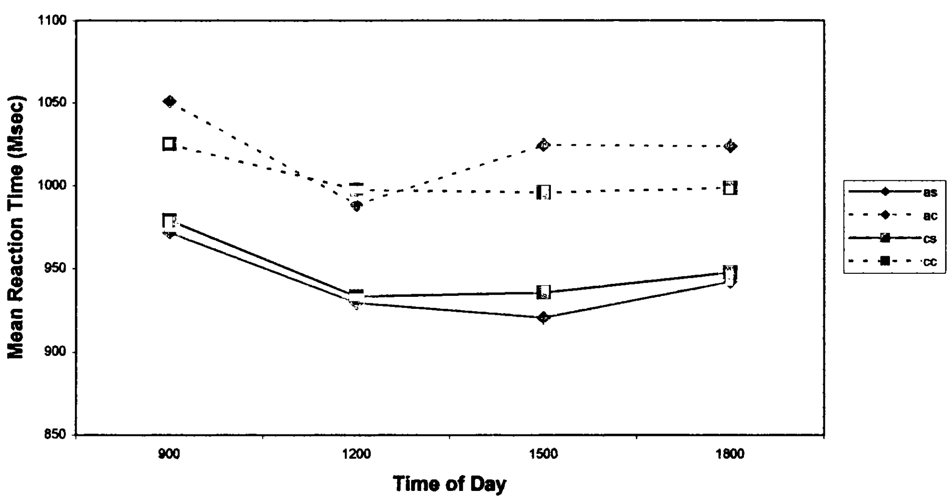


Figure 31(b): Mean Reaction Times for Each Icon Type at Each Time of Day in Experiment 5



6.4.2. Experiment 6

One difference remains between Experiments 1 and 2 and Experiments 3 and 5. In Experiments 1 and 2 a 'yes'/ 'no' response type was used, here the target icons were *not* always present in the distractor array and when this was the case a 'no' keyboard response was required while when the target was present a 'yes' keyboard response was needed. In Experiments 3 and 5 however, a mouse click response type was used where the target *was* always present in the distractor array and participants were required to click on the matching icon using the mouse. This is possibly an important difference between the cognitive demands of the experiments. Thus in Experiment 6 the 'yes'/ 'no' response type was reintroduced into the task used in Experiment 5. In every other way Experiment 6 was identical to Experiment 5. If it is the 'yes'/ 'no' response that is important the diurnal trend will closely resemble that seen for Experiments 1 and 2, if this is not the case however, the diurnal trend will differ from that seen in Experiments 1 and 2 and it can be concluded that it is icon distinctiveness that is important.

6.4.3. Conclusions

Experiment 5 has shown that visual memory alone is not of critical importance in determining diurnal performance trends and what must be borne in mind when attempting to improve the early morning performance of workers is that memory load must be quite high in order to improve performance at 0900.

Chapter 7

Experiment 6: The Effect of Response Type

7.1 Introduction

As noted in Chapter 6, in all the experiments previously reported there were a number of procedural details that were systematically varied. These were:

(i) *Visual Memory Component*

In Experiments 1, 2 and 5 the target icon disappeared from the screen before the search set appeared, requiring participants to remember the target icon when searching for it among the distractor array. Experiments 3 and 4 did not require participants to remember the target icons as these remained on screen during search among the distractor array.

(ii) *Difficulty of Response*

Experiments 1 and 2 used a 'yes' / 'no' response where participants were required to press one key if the target icon was present among the distractor array and another key if the target icon was absent. Experiments 3, 4 and 5 used a mouse click response where participants were required to click on the matching icon using the mouse thus the target icon was always present among the distractor array.

(iii) *Difficulty of Icon Discrimination*

Experiments 1 and 2 used a combination of multi-feature and gestalt, complex and simple, icons that did not systematically vary concreteness and complexity. Experiments 3, 4 and 5 used gestalt icons where the concreteness and complexity of the icons was orthogonally varied. These icons were thought to be more distinctive although this was not specifically measured, resulting in these icons being easier to discriminate.

It is interesting to note that each experiment showed different diurnal patterns in accordance with the above task demands (see Figure 32). The time-of-day trends were:

(i) *Experiments 1&2*

Both Experiments 1 and 2 involved a 'yes'/ 'no' keyboard response and a visual memory component and showed a similar pattern of diurnal variation (see Figures 32a&b). Experiment 1 replicated McFadden and Tepas' (1997) study using icons specifically designed for the experiment. Responses were fastest early in the day (0900/1200) and were slowest at 1500. Experiment 2 replicated the complex (multi-feature) icon trials from Experiment 1 and also used the same icons presented in a gestalt format (see icons in Figure 11). Again, responses were fastest early in the day (0900/1200) and were slowest at 1500.

Experiments 3&5

In Experiments 3 and 5 the complexity and concreteness of icons presented to participants was systematically varied. In this instance a basic search task was used and participants simply had to mouse click on the icon matching the present target. Experiment 3 provided evidence that even a simple icon search task is subject to diurnal performance fluctuations although the pattern of results was different from Experiments 1 and 2 (see Figure 32c). Responses were generally *slowest* at 0900 while the time taken to respond at all other times of the day were all similar. Experiment 5 was very similar to Experiment 3 but reintroduced the visual memory component while retaining the mouse click response. This experiment also found responses to be *slowest* at 0900 while the time taken to respond at all other times of the day were all similar. Therefore, generally for both Experiments 3 and 5 the significant difference in responses lay between 0900 and the rest of the day, thus Experiments 3 and 5 are revealing the same pattern of performance. This suggests that it is the type of response required (or differences in icon distinctiveness) rather than the visual memory component that determines the pattern of diurnal variation.

Figure 32(a): Mean Reaction Times for Complex and Simple Icons at Each Time of Day in Experiment 1

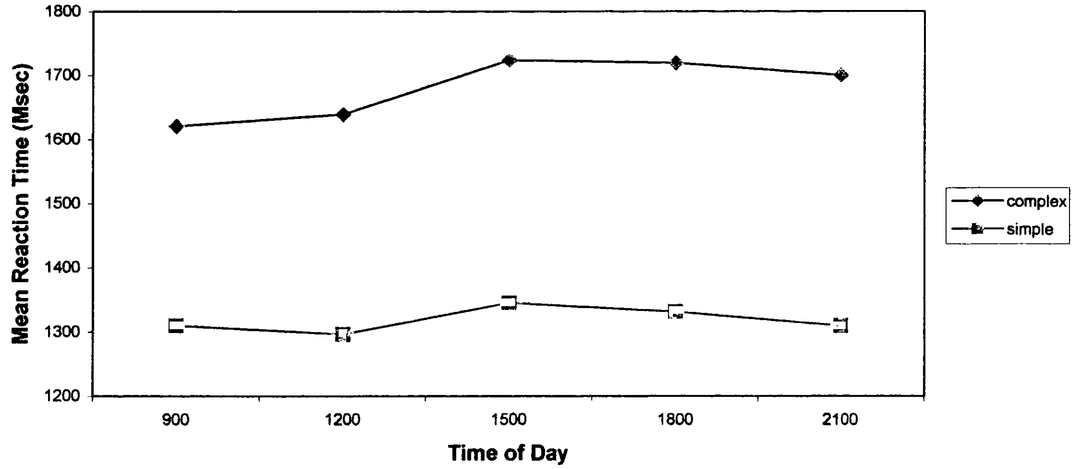


Figure 32(b): Mean Reaction Times at Each Time of Day for Multi-Feature and Gestalt Icons in Experiment 2

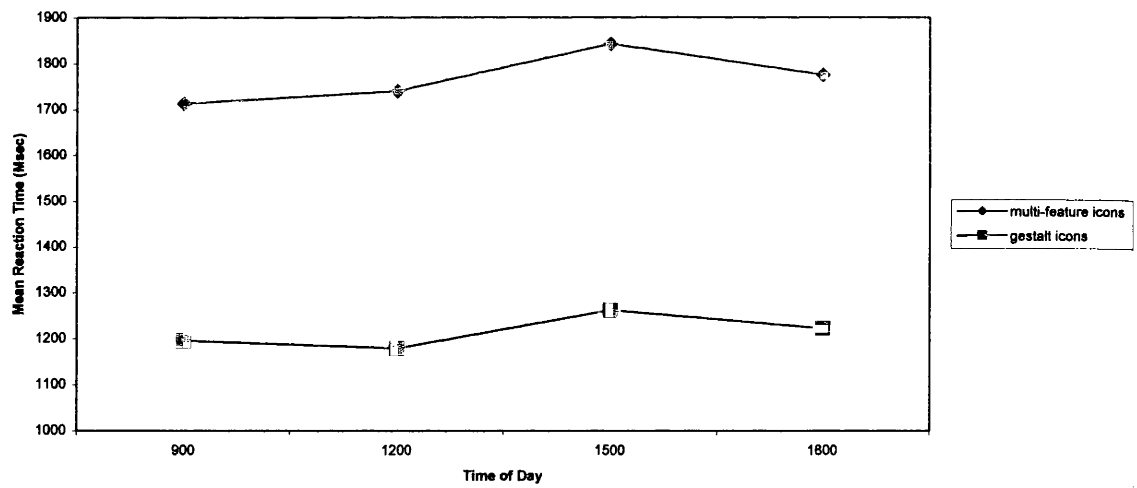


Figure 32(c): Mean Reaction Times for Each Icon Type at Each Time of Day in Experiment 3

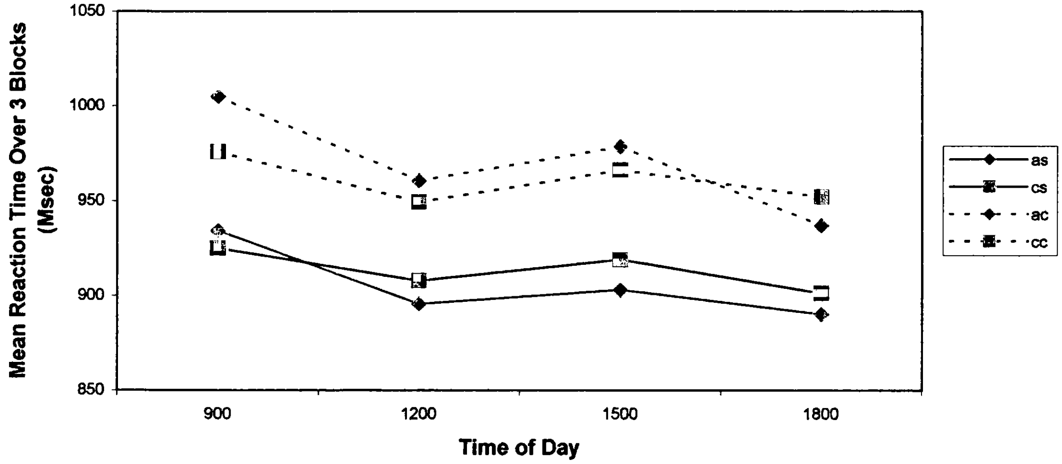
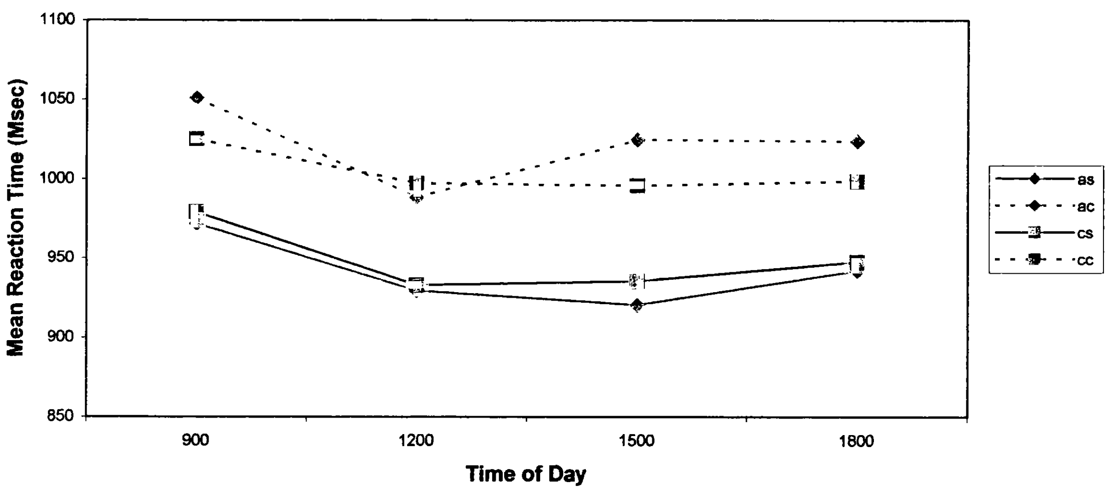


Figure 32(d): Mean Reaction Times for Each Icon Type at Each Time of Day in Experiment 5



Previous experiments have therefore shown that changes in task demands are important in determining the nature of the resultant diurnal trend. This is in accordance with prior research that has also demonstrated that time-of-day trends vary according to exact task demands (Smith, 1992). Consequently, the fact that the observed diurnal performance trends seen in Experiment 5 still differed from those seen in the first two experiments suggests that while it is possible that this is due to differences in the distinctiveness of the icons used, it is also possible that it is due to one fundamental difference that still remains in the exact demands of each task. This concerns response type. Experiments 1 and 2 are the only experiments to date to have used a 'yes'/'no' response.

In order to evaluate whether is the distinctiveness of the icons or the response type that is responsible for the differing diurnal trends, Experiment 6 will be identical to Experiment 5 in every respect apart from the response type, where the 'yes'/ 'no' response used in Experiments 1 and 2 will replace the mouse click response used in Experiments 3 – 5. Hence, Experiment 6 will use a visual memory component and a 'yes'/ 'no' response as used in Experiments 1 and 2. However, Experiment 6 will still use a more distinctive icon set than that used in Experiments 1 and 2.

7.1.1. Aims

Experiment 6 examined the following:

1) Whether the diurnal performance trends previously seen for each icon type in Experiment 5 were influenced by a change in task demands through the reintroduction of a 'yes'/ 'no' response type.
2) How the diurnal trend seen for Experiment 6 varied in comparison to the trend previously found for Experiment 5. If response type is the key to determining the diurnal trend observed then we might expect a pattern of performance more similar to that observed in Experiments 1 and 2. If not this would suggest that differences in icon distinctiveness are important.
3) Whether the temperature trend showed a relationship to performance trends.
4) As before, Experiment 6 also determined whether differences occur in the speed/accuracy of response for complex versus simple and abstract versus

concrete icons in a way similar to that seen in previous experiments.

It was expected that the exact time-of-day trend observed in Experiment 5 would again change as a consequence of further changes in the exact task demands, moreover it was expected that this change would result in a diurnal trend similar to that seen in Experiments 1 and 2. Due to the increased difficulty of response and therefore increased memory load involved in this experiment, a trend towards superior early morning performance was expected to reappear. Also a post-lunch dip in performance at 1500 was expected, as a decline in performance at this time was evident in Experiment 2 where visual memory and a 'yes'/ 'no' response type was used. As in all previous experiments, best performance was expected for simple icons, while icon concreteness was not expected to significantly influence performance.

7.2 Method

The paradigm used in Experiment 6 was identical to Experiment 5, with the exception that a 'yes'/ 'no' response was required instead of a mouse click.

7.2.1 Participants

Twenty-four undergraduates and postgraduates from the University of Wales Swansea participated. Nine were male and fifteen were female. The age range of participants was 18 to 25 years; the mean age was 21 years 3 months (standard deviation, 2 years 7 months). Some received course credit for their participation while others received a payment of £10.

7.2.2 Materials & Apparatus

The materials and apparatus used for Experiment 6 were identical to those used for Experiment 5.

7.2.3 Design

The counterbalancing of the experimental conditions across participants was identical to that in Experiment 5.

7.2.4 Procedure

To start the experiment, participants were required to click on an "ok" button present on the screen using the mouse. Once participants had clicked the "ok" button it turned grey and the target icon appeared on screen for a period of two seconds before disappearing. Immediately, the array of 9 distractor icons appeared on the screen. Participants were instructed to respond as quickly and as accurately as they could at all times.

Participants had to decide whether the target icon was present in the display or not. If the target icon was present, participants were instructed to press the 'Q' key, if the target was not present, participants were instructed to press the 'P' key. Each key was labeled 'Y' for 'yes' and 'N' for 'no' respectively to avoid confusion. Participants were asked to place their fingers on these keys for the duration of the experiment. Timing began when the icon array appeared and finished once participants had responded by pressing a key. The next trial began and the whole procedure repeated. The program ignored all erroneous key presses. No feedback was given regarding correctness of response. If no response was given within 6 seconds, the program moved on to the next trial automatically (see Figure 27 in Chapter 6 section 6.2.4 for example of procedure). This was repeated for each of the four blocks. There was a two-minute break between each block.

Eardrum temperature readings were taken before and after each test session and were taken from the right ear on every occasion. The procedure was identical at every test session.

Blocks of Trials and Randomisation of Icons

Selection of the targets and the distractor arrays was randomized so that each had an equal chance of being sampled and the participants saw a different set of stimuli during each test session. Four blocks of 108 trials were completed at each time-of-day. Each block of 108 trials used the same set of 72 icons. It was necessary to increase the number of blocks and trials to that used in Experiment 5, in order to gain a 2:1 ratio of present to absent icons as used in Experiments 1 and 2 while using the same icons used in Experiments 3 – 5. Seventy-two out of the 108 trials in each block were trials where the target icon was present among the distractors and each icon was presented once as a target. The remaining 36 trials were trials where the target icon was absent from the distractor array and one half of the 72 icons were used as targets in these absent trials. Note that in these 36 absent trials, half of the 72 icons were nominally the target icons, but of course these icons were not shown among the distractor array. In the next block of trials the other half of the 72 set of icons were used as targets in the absent trials. This was repeated for the remaining two blocks. Therefore, each icon was used three times as a target over two blocks of 216 trials but were only present among the distractor array twice. The array of 9 icons consisted of the target icon (when the target was present) and 8 distractors, or of 9 distractors (when the target was absent). The distractors in each present trial consisted of 2 icons from each of the four icon types (abstract-complex; abstract-simple; concrete-complex; concrete-simple). The distractors in each absent trial consisted of 2 icons from each of the four icon types plus an extra icon that was numerically the next icon on from the target used in the icon set. The location of these distractors within the array was randomized. Each icon appeared 26 times as a distractor over two blocks of 216 trials. Target icons from each of the icon conditions were presented 16 times at each of the 9 possible positions on the display over two blocks. Each position (1–9) on the display was used 24 times over two blocks.

7.3 Results

7.3.1. Response Times

Table 10 (a & b), illustrates a trend in performance that varied according to the time-of-day where generally response times were fastest at 0900 and/or 1200. Responses were clearly slowest at 1500 for all icon types. Performance was marginally superior at 0900 for the abstract-complex and concrete-complex icons but markedly better at 0900 for the abstract-simple icons. For the concrete-simple icon type performance was marginally superior at 1200. All icon types showed a marked fall at 1500 before showing good improvement at 1800. It can also be seen that response times were generally fastest in the simple conditions although this was not as consistent as in previous experiments. Concrete icons also failed to show any consistent reductions in response times.

A five-factor repeated measures analysis of variance was carried out. The factors were time-of-day (0900, 1200, 1500, 1800), icon complexity (complex versus simple), icon concreteness (concrete versus abstract), presence (icon present versus icon absent),

TOD	*Abstract-Simple Icons*							
	P/B1*	**P/B2**	**P/B3**	**P/B4**	**A/B1**	**A/B2**	**A/B3**	**A/B4**
0900	850.37 (184.99)	754.83 (176.18)	745.93 (134.14)	794.51 (174.25)	1396.49 (458.38)	1232.17 (364.60)	1167.79 (458.72)	1289.57 (291.60)
1200	809.55 (211.64)	782.68 (185.44)	818.76 (220.37)	817.99 (194.35)	1305.53 (408.69)	1353.51 (677.48)	1363.60 (527.73)	1384.40 (446.58)
1500	892.35 (142.36)	869.30 (238.85)	842.55 (141.39)	839.93 (194.71)	1391.36 (316.54)	1396.20 (408.26)	1411.73 (417.44)	1370.90 (455.73)
1800	791.18 (201.72)	799.88 (174.52)	806.71 (120.50)	784.06 (196.24)	1306.52 (461.07)	1267.64 (376.46)	1329.71 (354.54)	1309.13 (491.24)

TOD	*Abstract-Complex Icons*							
	P/B1	**P/B2**	**P/B3**	**P/B4**	**A/B1**	**A/B2**	**A/B3**	**A/B4**
0900	796.38 (160.54)	799.32 (181.06)	801.14 (180.84)	844.52 (149.05)	1292.99 (422.81)	1367.43 (391.27)	1365.33 (456.47)	1408.84 (444.42)
1200	796.83 (165.80)	802.04 (175.79)	1030.28 (1308.18)	803.69 (223.60)	1315.01 (357.32)	1308.10 (424.18)	1334.51 (436.94)	1319.42 (523.24)
1500	903.99 (130.42)	952.66 (189.24)	905.96 (133.82)	871.26 (187.69)	1289.11 (234.75)	1425.62 (410.50)	1405.05 (463.70)	1378.93 (451.78)
1800	825.37 (134.87)	864.80 (185.60)	844.29 (192.09)	790.80 (162.81)	1329.74 (414.14)	1326.24 (397.32)	1283.41 (338.81)	1413.81 (422.46)

Table 10(a): Mean (& standard deviations) Response Times for Present and Absent *Abstract Simple and Complex* Icons for Each Block at Each Time-of-Day

Continued…

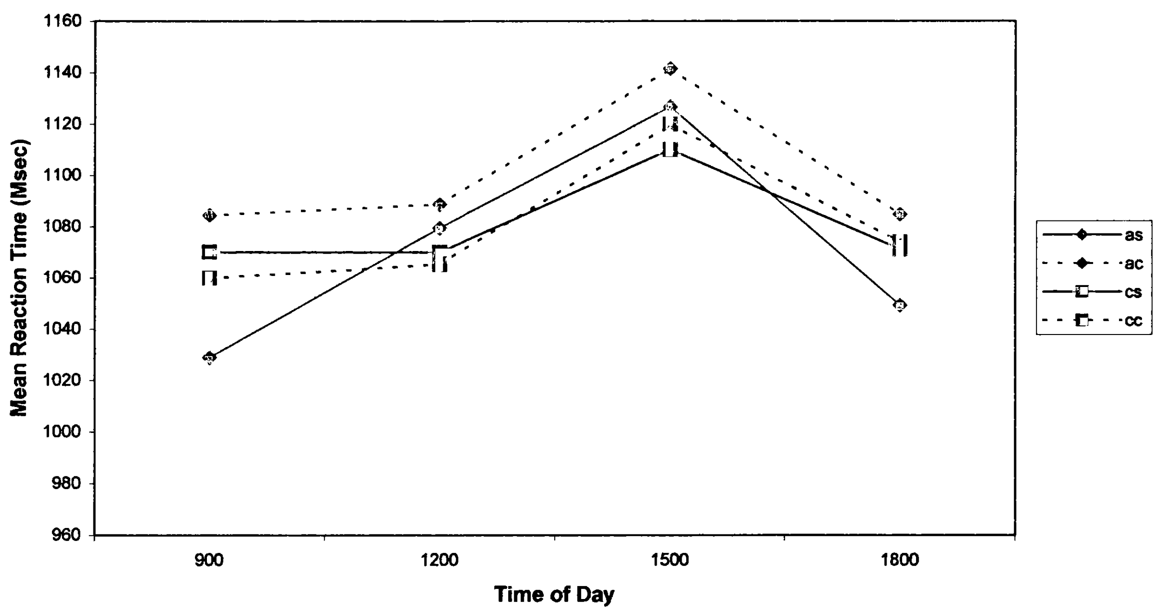
TOD	*Concrete-Simple Icons*							
	P/B1	**P/B2**	**P/B3**	**P/B4**	**A/B1**	**A/B2**	**A/B3**	**A/B4**
0900	796.93 (148.70)	779.21 (134.57)	795.44 (139.60)	808.07 (174.97)	1287.68 (357.19)	1324.67 (436.03)	1326.80 (337.67)	1442.57 (369.81)
1200	790.17 (188.36)	794.96 (187.75)	795.44 (200.54)	799.13 (163.21)	1302.29 (310.29)	1264.44 (387.02)	1296.18 (401.38)	1517.47 (522.20)
1500	844.10 (180.21)	834.02 (135.86)	824.32 (161.65)	803.70 (162.28)	1307.68 (297.35)	1416.41 (380.76)	1455.99 (453.87)	1394.45 (446.22)
1800	853.74 (197.87)	784.70 (145.24)	797.13 (142.95)	818.63 (178.91)	1283.33 (359.86)	1364.24 (486.33)	1359.90 (428.71)	1308.30 (384.69)

TOD	*Concrete-Complex Icons*							
	P/B1	**P/B2**	**P/B3**	**P/B4**	**A/B1**	**A/B2**	**A/B3**	**A/B4**
0900	790.40 (121.77)	810.91 (126.68)	807.93 (139.26)	798.68 (143.53)	1283.24 (359.45)	1330.55 (322.13)	1320.32 (315.53)	1336.80 (361.07)
1200	805.30 (151.77)	804.49 (163.03)	780.53 (164.04)	810.31 (200.98)	1325.99 (431.15)	1385.28 (393.17)	1298.07 (423.50)	1312.55 (442.16)
1500	841.25 (145.37)	846.79 (151.63)	837.89 (179.11)	855.22 (195.95)	1412.57 (483.19)	1399.51 (391.53)	1424.95 (427.35)	1341.91 (396.79)
1800	806.45 (191.06)	815.61 (191.14)	808.22 (174.12)	759.70 (120.76)	1328.27 (389.97)	1383.62 (424.72)	1378.96 (442.27)	1311.25 (356.56)

Table 10(b): Mean (& standard deviations) Response Times for Present and Absent *Concrete Simple and Complex* Icons for Each Block at Each Time-of-Day

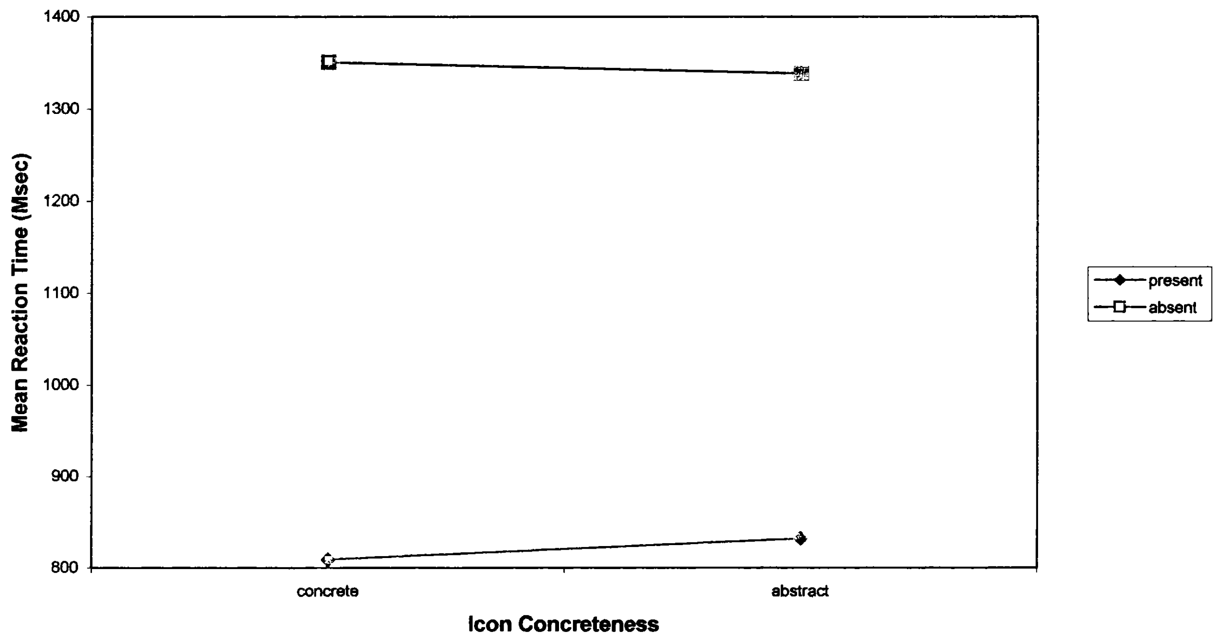
*P = Present; A = Absent; B1 = Block 1; B2 = Block 2; B3 = Block 3; B4 = Block 4 and block (1–4). See Appendix 5 for a full summary of results. The effect of time-of-day was significant ($F(3,69) = 2.99$, $p<0.05$). Response times were fastest when the target icon was present among the distractor array ($F(1,23) = 128.33$, $p<0.000$). Neither icon concreteness ($F(1,23) = 0.20$, p = ns) nor icon complexity ($F(1,23) = 1.24$, p = ns) significantly affected performance, nor did block number ($F(3,69) = 0.378$, p = ns) (see Appendix 5(a) for details of this analysis). From Figure 33 it can be seen that responses were fastest at 0900 and 1200 and slowest at 1500. In order to explore the time-of-day trends in further, Newman-Keuls analyses were carried out to identify exactly where significant differences lay. These analyses revealed that overall responses were significantly different between 0900 and 1500 and between 1200 and 1500. This suggests that time-of-day effects are most marked between early in the day (0900/1200) and the afternoon (1500).

Figure 33: Mean Reaction Times for Each Icon Type at Each Time of Day



A significant interaction was observed between icon concreteness and icon presence ($F(1,23) = 6.41$, $p<0.05$) on response times. Figure 34 shows response times decreased when abstract icons were absent from the distractor array but increased when abstract icons were present in the distractor array. From Figure 34, it can be seen that the difference in response times between present and absent icons was greater when the icons were concrete. Accordingly, simple main effects revealed that the difference in response times between present and absent icons was significant when icons were concrete ($F(1,23) = 7.85$, $p<0.01$) but not when they were abstract ($F(1,23) = 0.30$, p = ns). See Appendix 5(b) for details of this analysis.

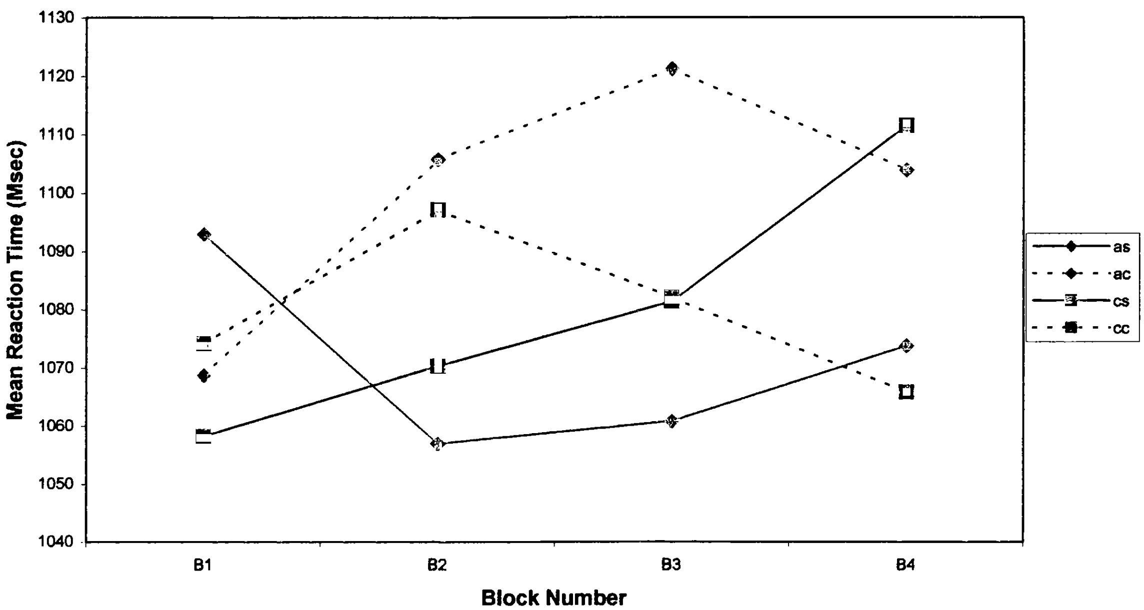
Figure 34: Interaction Between Icon Concreteness and Icon Presence



Another interaction was observed between icon concreteness, icon complexity and block number ($F(3,69) = 3.02$, $p<0.05$) on response times. From Figure 35 it can be seen that the difference between concrete and abstract icons generally increased as block numbers progressed. Simple simple main effects found the difference between concrete and abstract icons to be significant only for the complex icons at block 4 ($F(1,23) = 4.67$, $p<0.05$). There was also a trend for the difference between concrete and abstract icons to be significant only for simple icons at block 1 but this did not quite reach significance ($F(1,23) = 4.14$, $p = 0.054$). The difference between abstract and concrete simple icons at all other blocks was not significant: block 2 ($F(1,23) = 0.90$, p = ns); block 3 ($F(1,23)$ =

3.82, $p = ns$); block 4 ($F(1,23) = 1.07$, $p = ns$). Similarly, the difference between abstract and concrete complex icons at all other blocks was also not significant: block 1 ($F(1,23) = 1.07$, $p = ns$); block 2 ($F(1,23) = 0.43$, $p = ns$); block 3 ($F(1,23) = 0.35$, $p = ns$; see Appendix 5(b) for details of this analysis). The nature of this interaction is not clear and is probably the result of the nature of the data set used. No other interactions were significant.

Figure 35: Interaction Between Icon Concreteness, Icon Complexity and Block Number



7.3.2. Accuracy

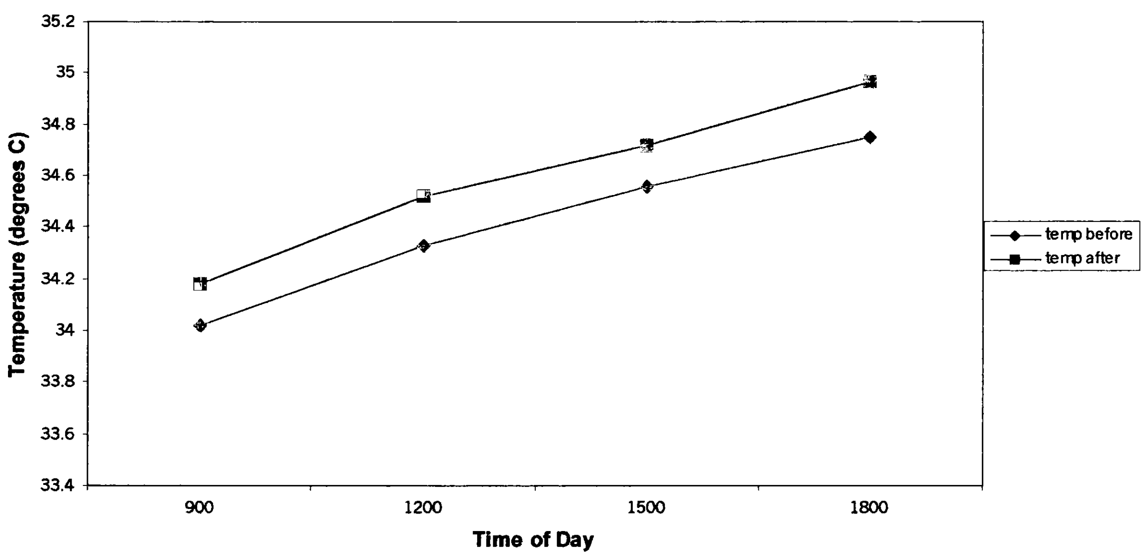
Once again, due to the nature of the task used in Experiment 6 there was very little variation in the accuracy data obtained. For full details of percentage accuracy rates see Appendix 5(c). Mean percentage accuracies for each participant ranged from 89.76% to 98.55% with the mode being 94.50% and 96.41%. Mean percentage accuracies for each time-of-day ranged from 95.77% at 0900 to 95.24% at 1800. Overall the mean percentage accuracy was 95.45%. Due to the presence of these ceiling effects no further analyses were conducted.

7.3.3. Temperature

A two factor repeated measures analysis of variance was carried out to examine the variation in temperature (before and after each time point) over the course of the day (0900, 1200, 1500, 1800). Significant differences in temperature values were found as a result of the time-of-day ($F(3,69) = 115.63$, $p<0.001$; see Figure 36). There was also a significant difference in temperature values before and after each testing session at each time-of-day ($F(1,23) = 78.69$, $p<0.001$) with temperatures being higher after each test session than they were before each session. No interactions were observed (see Appendix 5(a) for details of this analysis).

From Figure 36 it can be seen that, as previously, temperature gradually increased over the course of the waking day with the lowest temperature being seen at about 0900 and the highest at about 1800. Figure 36 also illustrates the increase in temperature at the end of each test session. To explore this time-of-day trend in temperature further, Newman-Keuls analyses were carried out to identify exactly where significant differences lay. These analyses revealed that temperature values differed significantly from one time of the day to the next.

Figure 36: Mean Temperature Values Before and After Each Time-of-Day



Comparison of the time-of-day trend seen for performance with that seen for temperature (see Figures 33 and 36) shows no evidence of a relationship between the two. Further, Newman-Keuls analyses showed significant differences in temperature that were not mirrored in performance.

7.4 Discussion

To summarise, response time varied in accordance with time-of-day. Responses were significantly faster at 0900 and 1200 than they were at 1500. Not surprisingly, response times were faster when the target icons were present in the distractor array. Participants' temperatures varied significantly over the day with temperature steadily increasing from early morning to early evening and there was no evidence of any relationship between temperature and performance.

7.4.1. Time-of-day Trends and Previous Research

Experiment 6 aimed to examine the effect of a 'yes'/ 'no' response type. The time-of-day trend observed differed markedly from that seen in Experiment 5. Comparison of Figure 33 (Experiment 6) with Figures 32a (Experiment 1) and 32b (Experiment 2) shows that the reintroduction of the 'yes'/ 'no' response type has increased the difficulty of the task in Experiment 6 to the extent that the trend towards superior early morning performance seen in Experiments 1 and 2 has reappeared. Further comparison of Figure 33 (Experiment 6) with Figures 32c (Experiment 3) and 32d (Experiment 5) demonstrates the performance decrements observed at 0900 in tasks where a mouse click response format is used. Clearly then, response type has an important role to play in terms of task demands and the resultant diurnal trends. Where the 'yes'/ 'no' response type has been used, in Experiments 1, 2 and 6 good performance has been observed early in the day at 0900, however, where the mouse click response type has been used in Experiments 3 – 5 performance was consistently poor at this time-of-day. Thus the reintroduction of the 'yes'/ 'no' response type in Experiment 6 has meant that the performance decrements at 0900 in Experiments 3 and 5 are no longer observed. The fact that the *general* pattern of findings in Experiments 1, 2 and 6 are the same suggests that icon distinctiveness does *not* determine the diurnal performance trend.

Finally, it is interesting to note that examination of Figure 33 again suggests that the diurnal trend seen may be more pronounced for abstract icons.

All icon types in Experiment 6 showed a significant fall in performance at 1500 but a significant improvement in responses after 1500 is not seen. Thus only Experiment 2 can provide support for the post-lunch dip, in that a significant improvement in responses is seen at 1800. This demonstrates that the post-lunch dip, like the early morning, is susceptible to changes in task demands and therefore is pliable (Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b; Smith et al, 1990; Smith, 1988).

Interestingly, Figures 32c&d show a clear difference in response time between complex and simple icons, however in Figure 33 this complex versus simple distinction disappears. The demands of a 'yes' / 'no' response appear to have removed this previously robust effect (see also McDougall et al, 1996, Byrne, 1993). Perhaps an explanation for the latter effect could be the distinctiveness of icons used in Experiment 6 which act to override the importance of icon complexity. Indeed distinctiveness has previously been found to override both complexity and concreteness effects under certain conditions (Arend et al, 1987; McDougall et al, 2000) and research has shown that even complex icons can be effectively searched for if they are distinctive enough (McDougall et al, 2000). So perhaps when icons are less distinctive, as in Experiment 2, icon complexity becomes important where the normal effectiveness of simple icons then becomes apparent.

So it has been clearly demonstrated that when the demands of the tasks used here are increased (through changes in visual memory, response type and icon discriminability), memory load is also increased and a high memory load removes early morning performance decrements observed in easier tasks. Furthermore, memory load does indeed need to be rather high in order for this effect to occur. Thus, once again clear support has been provided for the notion that the exact task demands are of vital importance in determining the exact nature of the diurnal performance trends observed (Smith 1992).

It may prove interesting to examine the effect of using the 'yes'/ 'no' response type alone, with no visual memory component, it is expected, that as visual memory appeared to have only a weak effect (see Experiment 5), removing this component would not influence the resultant time-of-day trend. Similarly, it would be interesting to examine the effect of using the icons used in Experiments 1 and 2 with/without a visual memory component and with a mouse click response type. On the basis of the findings reported here one might expect results similar to those reported for Experiments 3 and 5 as the experiments reported here have suggested that neither the visual memory component involved in a task nor the distinctiveness of icons is vitally important.

It seems fair to conclude that as the cognitive demands (memory load) of a relatively straight forward search task (that is, no meaning needs to be directly attached to an icon) increase then performance early in the day improves.

Diurnal Trends in Performance and the Arousal Model and Other Mechanisms

Again, one way of explaining the variation in performance that was observed in the present study was via the arousal theory framework. As we have seen, Experiment 6 supports the notion from the arousal theory that low arousal levels are best for high memory load tasks resulting in improved performance at 0900 (relative to tasks with a lower memory load). However, again performance failed to parallel the temperature rhythm and this is inconsistent with the arousal framework, although partial support is provided for Monk's (1982) theory that proposed that medium/high memory load tasks are mediated by an arousal rhythm that peaks three hours earlier (at 0900) than tasks using a low memory load task. Consequently, it is likely that more contemporary theories of diurnal performance variations may be more accurate. Again, it is likely that different cognitive subsystems have different circadian mechanisms controlling them, resulting in different diurnal performance trends (Folkard and Hill, 2002).

Diurnal Performance Trends and Temperature

Prior literature has reported very contradictory findings concerning the relationship between temperature and performance. Early literature reported a relationship between the two (for example, Kleitman, 1939, 1963) while later literature has failed to do so (for example, Owens et al, 2000). Here, once again no evidence of a relationship between performance and temperature was apparent.

7.4.2. Conclusions

Experiment 6 has confirmed that when considering time-of-day trends in icon tasks the exact demands of the task, not just the nature of the icons being utilized, are vitally important in assessing the risk of workers showing performance decrements according to the time-of-day. A difficult response type (combined with a visual memory component) removes performance decrements seen early in the day. Experiments 3, 5 and 6 all suggested that time-of-day performance trends may be more pronounced in the use of abstract icons.

Chapter 8

General Discussion

8.1. Aims of Thesis

The general aim of this thesis was to examine whether or not time-of-day effects influence an individuals' ability to complete tasks using icons. More specifically, this research examined whether different types of icon and specific task demands, in terms of visual memory, difficulty of response and difficulty of icon discrimination mediated the time-of-day trends seen. The effect of user experience and attaching meaning to an icon on observed diurnal performance trends was also considered briefly in Experiments 1 and 4. A further aim was to examine whether participants' diurnal temperature trends showed any relationship to the observed diurnal performance trends.

8.2. Summary of Results

Response times, but not accuracy of response, varied significantly in accordance with time-of-day for all experiments with the exception of Experiment 4, which did not show a significant diurnal performance trend. The rationale behind each experiment and the results will now be summarised and the primary conclusions that can be drawn from the experiments will be highlighted. Experiment 4 will be excluded initially and will be considered separately at the end. Task demands and findings for each experiment are summarised in Table 11.

Experiment 1 used icons designed to replicate McFadden and Tepas' (1997) study, visual memory was required and a 'yes'/ 'no' response type used. This study suggested that time-of-day effects in icon use reached significance when complex, but not simple, icons were used. Significant differences in response times lay between early/mid-morning (0900/1200) and the rest of the day (1500/1800/2100). Responses were fastest early in the day (0900/1200) and slowest at 1500 (see Figure 37a).

Experiment 2 replicated the complex icon trials in Experiment 1 and examined the effect of 'gestaltness' (the extent to which icons could be regarded as wholes or as a series of parts). As in Experiment 1, visual memory was required and a 'yes'/ 'no' response type was used. Responses were fastest early in the day (0900/1200) and slowest at 1500. Responses significantly improved at 1800 thus providing support for the post-

Experiment	*Response Type*	*Difficulty of Response*	*Icon Type*	*Difficulty of Icon Discrimination*	*Visual Memory Component*	*Resulting Overall Working Memory Load*	*Trend Towards Early Morning Peak Performance*	*Post-Lunch Dip*
1	'Yes'/ 'No'	High	Non-distinctive	High	Yes	High	Yes	No
2	'Yes'/ 'No'	High	Non-distinctive	High	Yes	High	Yes	Yes
3	Mouse	Low	Distinctive	Low	No	Low/None	No	No
4	Mouse	Low	Distinctive	Low	No	N/ASemantic	No	No
5	Mouse	Low	Distinctive	Low	Yes	Low	No	No
6	'Yes'/ 'No'	High	Distinctive	Low	Yes	High	Yes	No

Table 11: Variations in Task Demands and Resulting Changes in Peak Performance Trends in Icon Tasks

lunch dip. Response times for gestalt icons were also dramatically reduced relative to multi-feature icons from approximately 1700 milliseconds to 1200 milliseconds (see Figure 37b).

These initial experiments confirm that time-of-day effects do exist in icon tasks. Additionally, a slight, albeit not significant, trend towards superior performance at 0900 was observed for the icon types that resulted in increased memory load. It was proposed that simple icons improved discriminability and therefore reduced memory load as participants were only required to remember two pieces of information rather than the three pieces of information that had to be remembered for the complex icons. It was further proposed that according to the object-based theories of visual attention, the gestalt icons would be perceived as one object rather than three separate objects, consequently it was suggested that the gestalt icons would be remembered as one object rather than three (as would be the case for multi-feature icons) thereby again improving discriminability and reducing memory load. As the tasks used in Experiments 1 and 2 (and also Experiments 5 and 6) required participants to maintain information in memory during completion of the task, these tasks were thought to involve working memory. The slight performance peaks at 0900 (for the complex/multi-feature icons) were attributed to the high working memory load involved, while the slight performance peaks at 1200 (for the simple/gestalt icons) were attributed to the lower working memory load involved. Further, it can be concluded that gestalt icons markedly improve usability by reducing response times. To date, no one has shown that combining individual features into a gestalt can improve usability.

Experiment 3 used icons which varied in terms of their concreteness and complexity. In contrast to Experiments 1 and 2, there was no visual memory in the task and a mouse, rather than a 'yes'/ 'no' response was required. The icons used here (and in Experiments 4, 5 and 6) were considered to be more distinctive, and consequently more easily discriminated between, than those used in Experiments 1 and 2, although this was not explicitly measured. This experiment effectively tested time-of-day effects in a simple icon search task where participants could see the icon they were searching for at the side of the display and simply had to click with the mouse on the matching icon in the display, thus this task involved little, if any, working memory. Significant time-of-day effects were found and the trend appeared to be more pronounced for abstract icons. In general, the significant differences lay between 0900 and the rest of the day (see Figure 37c). Interestingly, in contrast to Experiments 1 and 2, performance was significantly slower at 0900 than at 1200. These performance decrements at 0900 are in sharp contrast to this being one of the *optimal* performance times in Experiments 1 and 2 (see Figures 37a&b). No support was provided for the post-lunch dip in performance.

The reasons why these contrasting patterns in the time-of-day effects appeared were examined in Experiment 5. Experiment 5 was identical to Experiment 3 with the exception that the visual memory component (where the icon to be searched for disappeared from screen) used in Experiments 1 and 2 was reintroduced, thereby testing the effect of visual memory on the observed time-of-day effect. Responses varied significantly according to the time-of-day, interestingly abstract-complex icons revealed the trend most similar to that seen previously and again it appeared that abstract icons may show more pronounced time-of-day effects (see Figure 37e). Response times were again significantly slower at 0900 in comparison to the rest of the day. Taken together the findings from Experiments 3 and 5 suggest that visual memory is not of critical importance in determining time-of-day performance trends. The performance decrements at 0900 were attributed to the working memory load of the task remaining low. No support was found for the post-lunch dip.

The remaining possibility was that response type influenced time-of-day trends. This was examined in Experiment 6. Experiment 6 was identical to Experiment 5 with the exception that the 'yes'/ 'no' response type used in Experiments 1 and 2 was reintroduced (in contrast to a mouse response in Experiments 3 and 5). As this experiment appeared to differ from Experiments 1 and 2 only in terms of icon distinctiveness, the influence of icon distinctiveness could also be considered. Response times significantly differed according to the time-of-day again it appeared that abstract icons showed the most variation in performance over the day. The significant differences lay between 0900 and 1500 and between 1200 and 1500 (see Figure 37f). Interestingly, the performance decrements seen at 0900 in Experiments 3 and 5 (see Figures 37c&e) were no longer present (see Figure 37f). This was attributed to the higher working memory component involved in this task, which was presumably a result of reintroducing the 'yes/ 'no' response type (this response not only required participants to establish if an icon was present or not resulting in the target having to be remembered for a longer period of time in the absent trials, but participants were also required to recode their 'yes' or 'no' response into a key press, this would increase memory load). Thus, Experiment 6 suggested that it was the 'yes'/ 'no' response type that was important in determining the diurnal performance trend seen. As the diurnal trend seen for Experiment 6 closely resembled that seen for Experiments 1 and 2, it was concluded that icon distinctiveness was *not* important in determining the resultant diurnal trend although these icon types do appear to be effective in reducing response times.

To summarise, the experiments outlined above have shown diurnal performance trends to vary according to the demands of the task. The result of the changes in task demands depended on how the task demand altered memory load and there were three aspects of each task that were considered to influence the resultant memory load (visual memory, difficulty of response, difficulty of icon discrimination). It appeared that these features do not necessarily need to be combined to exert an effect. For example, icon distinctiveness differed between Experiments 1 and 6 yet the general performance trend was the same, however whether difficulty of response exerts an effect independently of visual memory requires further investigation, although as visual memory had no effect in Experiment 5 it seems likely that response type does indeed operate independently. If working memory load is reduced performance decrements at 0900 were observed, while if working memory load is increased performance at 0900 improved. It appeared that it was the 'yes'/ 'no' response type that exerted the largest effects on the memory load involved in a task.

Experiment 4 appeared to differ from the other experiments and this was probably due to its' semantic memory component. Experiment 4 required participants to match a

Figure 37(a): Mean Reaction Times for Complex and Simple Icons at Each Time of Day for Experiment 1

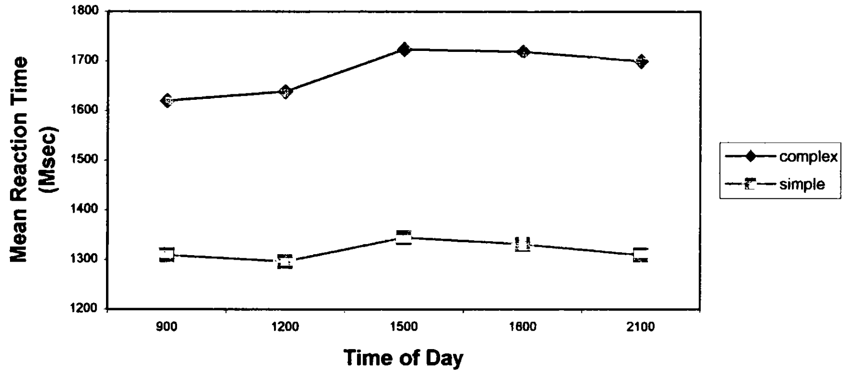


Figure 37(b): Mean Reaction Times at Each Time of Day for Multi-Feature and Gestalt Icons in Experiment 2

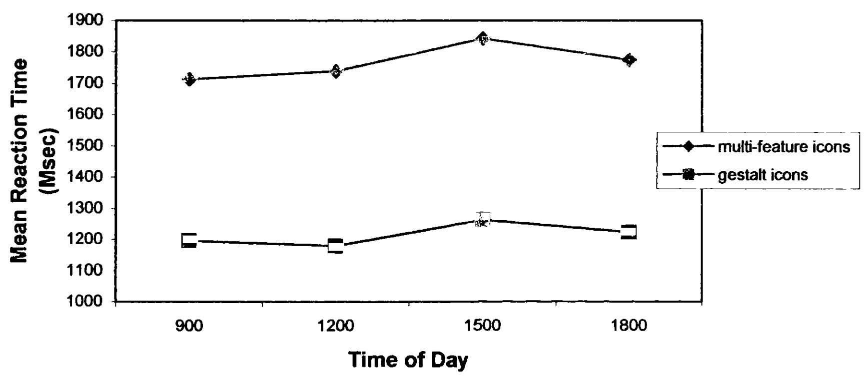


Figure 37(c): Mean Reaction Times for Each Icon Type at Each Time of Day for Experiment 3

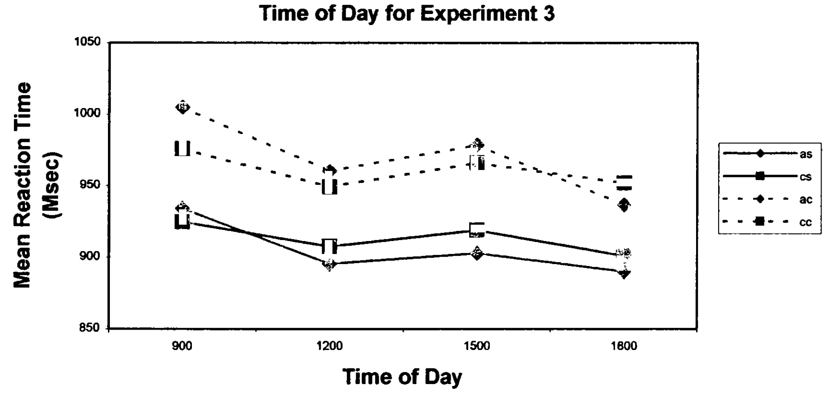


Figure 37(d): Mean Reaction Times at Each Time of Day for Each Icon Type for Experiment 4

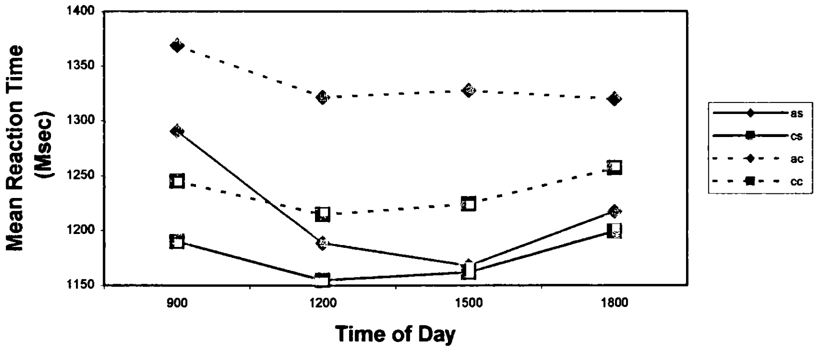


Figure 37(e): Mean Reaction Times for Each Icon Type at Each Time of Day for Experiment 5

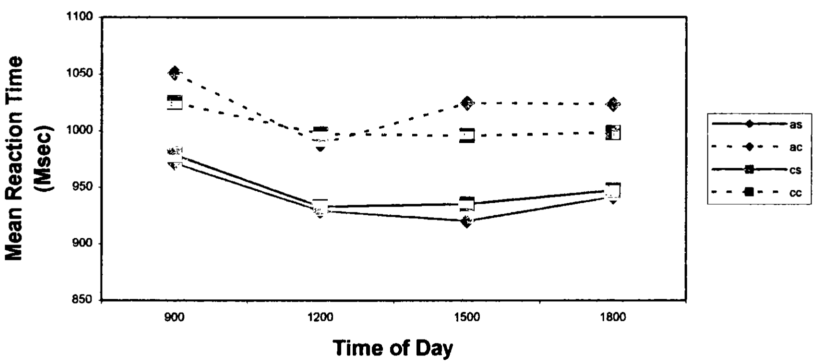
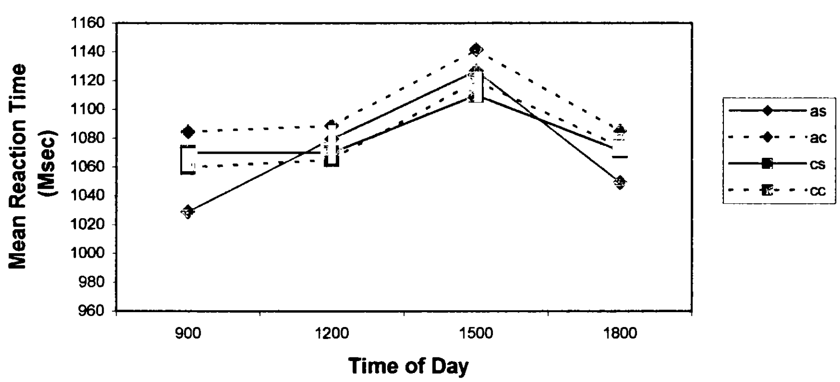


Figure 37(f): Mean Reaction Times for Each Icon Type at Each Time of Day for Experiment 6



function label with the appropriate icon. No visual memory component was required and a mouse click response type was used. This experiment examined the effect of directly attaching meaning to an icon. Responses did not significantly vary in accordance with the time-of-day, however the abstract-complex icon type showed a pattern similar to that seen in other experiments (see Figure 37d). Performance was significantly faster for simple icons and also for concrete icons, thus icon concreteness had an effect. Consequently, there is evidence that this semantic memory task was not susceptible to time-of-day effects. Clearly, time-of-day effects in semantic memory icon tasks need to be explored further.

8.3. Primary Conclusions

Five main conclusions can be drawn from the above findings:

1) Time-of-day effects in icon tasks do exist and abstract icons may be the most susceptible.
2) Time-of-day effects emerge even in a simple search task, but the exact demands of the task are important in determining the exact diurnal performance trend seen: (a) visual memory is *not* of critical importance in the emergence of time-of-day effects; (b) a 'yes'/ 'no' response type *is* important in determining the exact diurnal trend seen; (c) icon distinctiveness is *not* important in determining the time-of-day trend; (d) tasks involving semantic memory are *not* susceptible to time-of-day effects in the task used here.
3) The above changes in task demands seem to exert their effects through memory load: these effects seen can be explained with reference to previous memory and time-of-day research.
4) Gestalt icons markedly improve usability, as do generally simple icons. Concrete icons also improve usability where meaning is important.
5) Diurnal temperature trends show no obvious relationship to diurnal performance trends.

Each of these conclusions will now be considered in turn.

8.3.1. Time-of-Day Effects in Icon Tasks

All experiments, with the exception of Experiment 4, demonstrated time-of-day effects in icon use. Moreover, although Experiment 1 suggested that this was only the case for complex icons, subsequent experiments (for example, Experiments 3, 5 and 6) showed that both complex and simple icons were susceptible, although abstract icons seemed to show more pronounced diurnal trends. However, all icons appear to have the potential to reveal time-of-day effects and as stated in Experiment 3, the evidence for abstract icons showing more pronounced time-of-day effects is inconclusive but the emerging pattern appeared noteworthy.

8.3.2. Time-of-Day Effects and Changes in Task Demands

The difference in the performance trends between experiments could be attributed to one of three things: visual memory; difficulty of response; difficulty of icon discrimination (see Table 17 for a summary). These factors were varied over six experiments, section 8.2 outlines the task demands of each experiment and the resulting time-of-day trends. Briefly, the post-lunch dip was shown to be flexible (Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b; Smith et al, 1990; Smith, 1988) and (Experiment 4 aside) the main way in which the diurnal performance trends in each experiment differed concerned early morning performance at the 0900 time point. As will be seen in the next section, this can be explained in terms of previous research into time-of-day and memory.

8.3.3. The Memory Load Framework

With the exception of Experiments 3 and 4, all experiments required the participant to remember the icon being searched for while completing the task. Information had to be held in memory while a response was generated. This had to be done for three components of the task, which varied between experiments. Firstly, in experiments where the target icon disappeared from screen, visual memory was required to remember this image until a response had been made. Clearly, having to remember the target icon automatically increases the memory load of the task. Visual memory was not needed in Experiments 3 and 4 where the target icon (or function label) remained on screen while participants responded, thus working memory load was lighter in these experiments. Secondly, in experiments where a 'yes'/ 'no' response was required the 'yes'/ 'no' answer had to be recoded into a key press (the participant had to remember which key to press to give their answer) and the participant also had to decide whether the target icon was present or not. In other experiments, where a mouse click response was required participants were asked to simply click on the matching icon that was always present among the distractor array and no recoding of responses into 'yes'/ 'no' key presses was necessary. Consequently experiments using a mouse click required a lighter memory load than those requiring a 'yes'/ 'no' response. Finally, it was proposed that as distinctive icons improve discriminability, this meant that target icons would be identified more rapidly when a distinctive icon set was used thus the participant would not be required to memorise the target for the same period of time as with a non-distinctive icon set, in this way increased icon distinctiveness was also believed to reduce memory load.

Tasks that showed severe performance decrements at 0900 involved no visual memory, used a mouse click response type and distinctive icons. These tasks seem likely to involve a lighter memory load (Experiments 3 and 5). Conversely, tasks that showed improved or even superior performance at 0900 involved a visual memory requirement and a 'yes'/ 'no' response type (Experiments 1, 2 and 6). Whether the icons were distinctive or not does *not* appear to be of importance for this effect. For instance, Experiments 1 and 2 used non-distinctive icons while Experiment 6 used distinctive icons and both showed improved performance at 0900. It is fair to conclude that tasks involving both, a visual memory component and a 'yes/ 'no' response, are higher memory load tasks. Further, as discussed earlier, it seems response type probably acts independently of visual memory.

It seems reasonable to suppose that the above differences in the working memory load involved in the completion of the task account for the different diurnal performance trends seen. In support of this notion, the fact that performance improved at 0900 as the memory load of a working memory task increased in Experiments 1, 2 and 6, is consistent with literature finding the peak in working memory performance to occur earlier on more highly loaded memory tasks (Folkard and Monk, 1980; Folkard, 1983).

Only Experiment 2 provided support for the post-lunch dip (Kleitman, 1939, 1963; Owens et al, 2000; Lenne et al, 1997), the remaining experiments suggested that the post-lunch dip could be manipulated by changing task demands. Thus consistent with previous work, the post-lunch dip maybe pliable (Craig et al, 1981; Smith and Miles, 1986a, 1986b, 1987b; Smith et al, 1990; Smith, 1988).

Experiment 4 did not demonstrate significant differences in response times according to changes in the time-of-day, although it is interesting to note that the abstract-complex icon type again revealed a pattern similar to that seen previously. However, the task used in Experiment 4 required different cognitive subsystems to complete the task than the other experiments. More specifically, the task used in Experiment 4 required the use of semantic memory, where participants were required to remember the meaning of each icon. After 0900 performance improved over the day before generally beginning to decline again at 1800. This pattern of performance is therefore inconsistent with research that has shown performance of semantic memory tasks to improve later in the day (Millar et al, 1980; Tilley and Warren, 1983; Smith, 1987a). Research has also shown that changing the nature of a semantic memory task can diminish any existing time-of-day effect (Smith, 1987a), suggesting that it is some specific aspect of the task used for Experiment 4 that resulted in no time-of-day effect.

The fact that the exact diurnal performance trends observed vary according to exact task demands is supported by previous work and illustrates the fact that the whole testing situation must be considered when interpreting diurnal performance trends (Smith, 1992). Support has also been provided for Folkard and Hills' (2002) statement that working memory tasks use different cognitive subsystems and the diurnal pattern observed is an "outcome of a combination of different trends associated with the different cognitive mechanisms involved" (p.57).

The arousal theory framework that was developed initially by Colquhoun (1971), was long used as an explanation for such variations in performance according to the memory load of the task, this framework theorized that the optimum level of arousal is high for low memory tasks and that a low level of arousal is optimum for high memory load tasks. The results presented here are certainly consistent with this. However, an important point to note is that the arousal theory is so general it is extremely difficult to falsify (Folkard, 1983), further a central point of the arousal framework was that performance trends should be related to temperature trends, as will be seen in section 8.3.5, this is certainly not supported by the research presented here. Consequently, it seems that more contemporary theories (for example, Monk and Leng, 1982; Monk et al, 1983; Monk et al, 1989; Dijk et al, 1992; Johnson et al, 1992; Folkard and Akerstedt, 1992) may be more accurate. However, the research reported here cannot suggest what the mechanisms underlying the observed time-of-day effects may be and indeed this was not the intention, the aim of this work being to establish if time-of-day effects in icon usability exist, and how these effects may be different for different icons or under different conditions.

8.3.4. The Effects of Icon Type on Performance

Experiment 2 compared performance on icons that had their features presented separately (multi-feature icons) with icons that had the same features combined to form one whole, or a gestalt (gestalt icons). It was found that gestalt icons reduce response times by approximately 500 milliseconds, a finding that is vitally important to the design of icons for use in time-critical applications. Furthermore, icons that were both distinctive and gestalt (Experiments 3-6) showed slightly faster response times than icons that were non-distinctive gestalts (Experiment 2). As speed of response is a vital component of many tasks using icons (for example, air traffic control), only time-of-day effects for the gestalt icon format was considered in latter experiments. Experiment 3-6 also show that generally simple icons improve usability, as do concrete icons where meaning is important, and this is consistent with much previous research (Byrne, 1993; McDougall et al, 1996; Stammers, 1990; Stammers et al, 1989; Stotts, 1998; Blankenberger and Hahn, 1991).

8.3.5. Diurnal Performance Trends and Temperature

Overall, no relationship was apparent between the diurnal performance trends and the diurnal temperature trends, supporting research that has also failed to find any relationship (for example, Owens et al, 2000). Consequently, the research reported here supports Carrier and Monks' (2000) suggestion that performance should be considered independently of physiological rhythms.

8.4. Practical Implications

It would appear that icon tasks are equally as sensitive to time-of-day effects as other tasks such as the letter cancellation task (LCT) (for example, Casagrande et al, 1997) and reading comprehension (for example, Englund, 1979), probably because symbology tasks use the same underlying cognitive operations as previously tested tasks using working memory.

Consequently, several practical implications can be drawn from the experiments reported. Firstly it is clear that in order to improve response times in time-critical applications gestalt icons should be used. Similarly, simple icons would also enhance usability in applications where speed of response is important, as would concrete icons in situations where meaning is a factor. If superior performance is required earlier in the day, the memory load of the task needs to be increased, note however that the memory load of the task must be quite high to achieve this, introducing a visual memory component alone is not enough. In tasks where it would be advantageous to minimize the magnitude of diurnal performance variations, it seems concrete icons may be advantageous. Alternatively, to be free of diurnal performance fluctuations consider using a semantic memory task but beware that changing aspects of this task may produce significant diurnal performance fluctuations.

Folkard's (1983) conclusions that the best time-of-day to perform a task depends on the nature of the task itself and that "it is undoubtedly the case that such (time-of-day) effects exist, that they are relatively unavoidable and that they have important practical, as well as theoretical, implications for the study of human performance efficiency" (p.268), clearly also hold for performance on icon tasks.

8.5. Future Research

The research reported here has provided a broad overview of the time-of-day trends that can be expected to emerge in icon tasks under differing task demands and Table 11 provides a framework for the explanation of time-of-day effects in icon tasks. However, this framework needs further exploration in order to pin down the exact time-of-day trends that can be expected under certain conditions, thus future work needs to consider further the role of different icon characteristics and other changes in task demands.

From the experiments reported here it would appear that difficulty of response may be a very important feature of a task that may alter memory load to such an extent that the performance decrements usually seen at 0900, when using a mouse click response, disappear. Future experiments should verify this and need to determine whether visual memory is an important part of this effect. To expand, the findings in Experiment 6 cannot be directly attributable to response type alone but to the combination of visual memory and response type, Experiment 6 did not separate these effects. Consequently, another experiment would be useful using the same paradigm as Experiment 6 but with the visual memory component removed, from the results of Experiment 5, which showed visual memory to be non-critical, it is expected that the results of such research would validate the importance of difficulty of response on the resultant diurnal performance trend. The effect of changing task demands, for example the effect of using a 'yes'/ 'no' response and the effect of visual memory, in a semantic memory icon task also need to be systematically examined.

The task used in Experiment 4 was similar to the process that a learner driver undertakes, icons may appear to be quite obscure at first but the meaning of these icons is rapidly learned. From the Experiment 4 task it would appear that learner drivers would be free from diurnal performance variations, however this does not account for the performance of an experienced driver where the meaning of the icons was well learned and so would not be such an important issue. Future experiments may wish to examine this factor. Indeed it would be useful to examine time-of-day effects in the use of existing road signs, this was not considered here as the aim was to develop icons that closely paralleled those used by McFadden and Tepas (1997). Further, it would be interesting to examine this with groups of participants who differed in their level of driving experience (for example, learners versus taxi drivers). Similar experiments could use pilots and cockpit displays, air traffic controllers and their displays and chemical processing plant workers and their interfaces.

Once firm confirmation of the nature of any existing diurnal performance trends were found attention could then move to ways of eliminating the performance fluctuations. From the studies reported here it appears that time-of-day trends emerge in one form or another regardless of changes in task demands, thus, with the exception of semantic memory, we have shown that although the exact trend can be manipulated, it does not appear that diurnal performance fluctuations can be eradicated by changing task demands. Consequently, we may need to look elsewhere for a solution, perhaps the use of a psychostimulant such as caffeine would be worthy of investigation. Furthermore, future experiments could examine the influence of individual differences on peak performance trends, for instance future work could investigate the effect of sleep deprivation on any observed time-of-day effects in an attempt to establish if this exacerbates the problem. Also, individual differences in circadian rhythms such as the morningness of the subject are known to influence peak performance times (Monk and Leng, 1986), consequently, the effect of morningness on performance of an icon task warrants further examination. As we have seen here, task demands are also vitally important in establishing diurnal performance trends, thus both task demands and personal characteristics of the subject must be considered (Owens et al, 2000) before definite conclusions can be drawn about when certain tasks are best carried out.

Further, the research reported here concerned performance trends observed during the *day*. The outcome of research that examined performance on icon tasks into the night would have much practical significance as many tasks requiring the use of icons are carried out over the whole 24 hour period for example, long distance lorry driving or monitoring chemical processing plants. Interestingly, Carrier and Monk (2000) state that when "the sleep-wake cycle is suspended and data collection in extended into the night, circadian performance rhythms appear generally to be predictable from the circadian temperature rhythm" (p.722), future work could therefore look at predicting peaks and troughs in performance during the night time from temperature readings, if this was found to be reliable the practical significance would be immense.

Although it appears that the changes in task demands, and the different diurnal performance trends resulting from these changes, can be explained in terms of memory load, future work needs to examine this more closely. For instance, icon discriminability is associated with search and while it is plausible that this would impact on memory load (for reasons discussed in previous chapters) future work should attempt to determine to what extent the effect of icon discriminability is attributable to memory and to what extent it is attributable to search. Icon discriminability is clearly a factor in search but how much this impacts on memory load has not been firmly established. It needs to be established whether search and memory load are indeed one and the same.

One factor that was believed to influence discriminability was the distinctiveness of the icon set. Although the experiments reported here and previous research has shown that icon distinctiveness influences response times (Aspillaga, 1996; Fisher and Tanner, 1992; Byrne, 1993), as shown in Experiment 6, icon distinctiveness does *not* appear to play an important role in time-of-day trends. That this is indeed the case requires further verification using a study where distinctiveness is measured and systematically varied. Furthermore, Experiments 3, 5 and 6 suggested that abstract icons may show more pronounced time-of-day performance trends, this clearly needs to be investigated further. Moreover, most of the results reported are explained in terms of changes in memory load, however the possibility that abstract icons show more pronounced time-of-day effects cannot be explained in terms of memory load. Perhaps the clarity of an icon's meaning is important, this could be explained in terms of semantic memory and when a task allows clear understanding of an icons' meaning no diurnal performance trend emerges. Indeed the results of the semantic memory task in Experiment 4 that showed no time-of-day effect, can provide support for this explanation. Future work needs to examine this more closely. Additionally, research has shown that articulatory distance influences response times when icon screen positions were randomized but not when the icon screen positions were fixed (Blankenberger and Hahn, 1991), further, the difference in response times for abstract and concrete icons has been found to be greatly reduced in arrays where icon positions were fixed (Green and Barnard, 1990). Consequently, it follows that if the experiments reported here used fixed screen positions rather than random ones, then the reported time-of-day trends may alter or even diminish, a factor that is clearly worthy of investigation in future research.

Such research could culminate in the proposal of safety guidelines for consideration by services where safety is of critical importance, for example, by airlines. Further, such work may led to proposals to improve performance on tasks where safety may not be important but achievement is, for instance suggestions could be made for the best time-of-day to teach computer orientated subjects in schools. This may become a very significant factor in the future of schools where the use of information and communications technology is rapidly advancing.

References

Akerstedt.T. (1977). "Inversion of the sleep-wakefulness pattern: effects on the circadian variations in psychological activation," Ergonomics, 20, 459-474.

Akerstedt.T., Levi.L. (1978). "Circadian rhythms in the secretion of cortisol, adrenaline and noradrenaline," European Journal of Clinical Investigation, 8, 57-58.

Altabet.S.C. (1995). "Intelligence," Dissertation Abstracts Journal: Section B: The Sciences and Engineering, 55 (9-B), 4099.

Anderson.K.J., Revelle.W. (1994). "Impulsivity and time of day: is rate of change in arousal a function of impulsivity?" Journal of Personality and Social Psychology, 67(2), 334-344.

Arend.A., Muthig.K., Wandmacher.J. (1987). "Evidence for global superiority in menu selection by icons," Behaviour and Information Technology, 6, 411-426.

Aschoff.J., Wever.R.A. (1962). "Spontanperiodik des menschen bei ausschluss aller zeitgeber," Naturwissenschaften, 49, 337-342.

Aschoff.J. (1965). "Circadian rhythms in man: a self-sustaining oscillator with an inherent frequency underlies human 24-hour periodicity," Science, 148, 1427-32.

Aschoff.J., Wever.R. (1981). "The circadian system of man." IN J.Aschoff. Handbook of Behavioural Neurobiology, Vol. 4, Biological Rhythms. New York: Plenum Press.

Aspillaga.M. (1996). "Perceptual foundations in the design of visual displays," Computers in Human Behaviour, 12, 587-600.

Baddeley.A.D., Hatter.J.E., Scott.D., Snashnall.A. (1970). "Memory and time of day," Quarterly Journal of Experimental Psychology, 22, 605-609.

Banks.W.P., Flora.J. (1977). "Semantic and perceptual processes in symbolic comparisons," Journal of Experimental Psychology: Human Perception and Performance, 3, 278-290.

Barker.P.G., Najah.M., Manji.K.A. (1987). "Pictorial communication with computers," International Journal of Man-Machine Studies, 27, 315-336.

Bateson.A.G., Alexander.R.A., Murphy.M.D. (1980). "Cognitive processing differences between novice and expert computer programmers," International Journal of Man-Machine Studies, 26, 649-660.

Berger.D.D.S. (2000). "The effects of time of day of task presentation and individual differences in self-reported optimal performance periods on the memory of adolescents enrolled in an early starting time high school," Dissertation Abstracts International: Section A: Humanities and Social Sciences, 61(4-A), 1287.

Blake.M.J.F. (1967a). "Time of day effects on performance in a range of tasks," Psychonomic Science, 9, 349-350.

Blake.M.J.F. (1967b). "Time of day effects on performance in a range of tasks," Psychonomic Science, 9, 345-350.

Blake.M.J.F. (1971). "Temperament and time of day." IN W.P. Colquhoun (Ed). Biological Rhythms and Human Performance. London: Academic Press.

Blake. M.J.F., Corcoran.D.W.J. (1972). "Introversion-extraversion and circadian rhythms." IN P.Colquhoun (Ed). Aspects of Human Efficiency. London: English Universities Press.

Blake.M.J.F. (1976). "Relationship between circadian rhythm of body temperature and introversion-extraversion," Nature, 215, 896-897.

Blankenberger.S., Hahn.K. (1991). "Effects of icon design on human-computer interaction," International Journal of Man-Machine Studies, 35, 363-377.

Boersema.T., Zwaga.H.J.G. (1996). "Pop-out in visual search requires practice," Perceptual and Motor Skills, 82, 355-365.

Bonnet.M.H., Arand.D.L. (1994). "The use of prophylactic naps and caffeine to maintain performance during a continuous operation," Ergonomics, 37(6), 1009-1020.

Borbely.A.A. (1982). "A two process model of sleep regulation," Human Neurobiology, 1, 195-204.

Breen-Lewis.K., Wilding.J. (1984). "Noise, time-of-day and test expectations in recall and recognition," British Journal of Psychology, 75, 51-63.

Broadbent.D.E. (1971). "Decision and Stress." London: Academic Press.

Broughton.R.J. (1998). "SCN controlled circadian arousal and the afternoon 'nap zone'," Sleep Research Online, 1, 166-178.

Budzynski.C.A., Bingman.V.P. (1999). "Time of day discriminative learning in homing pigeons, Columbia Livia," Animal Learning and Behaviour, 27(3), 295-302.

Byrne.M.D. (1993). "Using icons to find documents: simplicity is crticial," Proceedings of INTERCHI '93, 446-453.

Campbell.S.S. (1984). "Duration and placement of sleep in a 'disentrained' environment," Psychophysiology, 21 (1), 106-113.

Campbell.S.S., Zulley.J. (1988). "Napping as a biological rhythm: disentrainment of the human sleep/wake system." IN W.P. Koella., F.Obal., H.Schulz (Eds). Sleep'86. New York: Gustav Fischer.

Campbell.S.S. (1992). "Effects of sleep and circadian rhythms on performance." IN Smith.A.P., Jones.D.M. (Eds). Handbook of Human Performance Vol.3. London: Academic Press.

Carlson.N.R. (1995). "Foundations of physiological psychology," (3rd ed). USA, Allyn & Bacon.

Carrier.J., Monk.T.H. (2000). "Circadian rhythms of performance: new trends," Chronobiology International, 17 (6), 719-732.

Casagrande.M., Violani.C., Curcio.G., Bertini.M. (1997). "Assessing vigilance through a brief pencil and paper letter cancellation task (LCT): Effects of one night of sleep deprivation and of the time of day," Ergonomics, 40(6), 613-630.

Casagrande.M., Curcio.G., Tricarico.M.,Ferrara.M., Porcu.S., Bertini.M. (2000). "A letter cancellation task (LCT) for assessing diurnal variations in vigilance: effects of attention and task duration," Giornale Italiano di Psicologia, 27 (2), 401-428.

Chase.W.G., Simon.H.A. (1973). "Perception in chess," Cognitive Psychology, 4, 55-81.

Chelminski.I. (2000). "Modulating effects of time of day on age differences in memory performance: Attention versus inhibition," Dissertation Abstracts International – Section B – The Sciences and Engineering, 60 (11-B), 5801.

Chi.M.T.H., Glaser.R., Rees.E. (1983). "Expertise in problem solving." IN Sternberg.R. (Ed.), Advances in the Psychology of Human Intelligence, Vol. 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp.17-76.

Chiles.W.D., Alluisi.E.A., Adams.O. (1968). "Work schedules and performance during confinement," Human Factors, 10, 143-196.

Christie.M.J., McBrearty.E.M.T. (1979). "Psychophysiological investigations of post lunch state in male and female subjects," Ergonomics, 22, 307-325.

Clements.P.R., Hayer.M.D., Vermillion.M.E. (1976). "Psychometric, diurnal and electrophysiological correlates of activation," Journal of Personality and Social Psychology, 33, 387-394.

Coltheart.M., (1980). "Iconic memory and visible persistence," Perception and Psychophysics, 27, 183-228.

Coltheart.M., Rastle.K. (1994). "Serial processing in reading aloud: evidence for dual-route models of reading. Special section: modelling visual word recognition". Journal of Experimental Psychology: Human Perception and Performance, 20, 1197-1211.

Colquhoun.W.P. (1960). "Temperament, inspection efficiency and time of day," Ergonomics, 3, 377-378.

Colquhoun.W.P., Corcoran.D.W.J. (1964). "The effects of time of day and social isolation on the relationship between temperament and performance," British Journal of Social and Clinical Psychology, 3, 226-231.

Colquhoun.W.P., Blake.M.J.F., Edwards.R.S. (1968a). "Experimental studies of shiftwork I: a comparison of 'rotating' and 'stabilised' 4 hour shift systems," Ergonomics, 11, 437-453.

Colquhoun.W.P., Blake.M.J.F., Edwards.R.S. (1968b). "Experimental studies of shiftwork II: stabilised 8 hour shift systems," Ergonomics, 11, 527-546.

Colquhoun.W.P. (1971). "Biological rhythms and human performance." London, Academic Press.

Colquhoun.W.P. (1972). "Aspects of human efficiency: diurnal rhythm and loss of sleep." London: English Universities Press.

Colquhoun.W.P. (1981). "Rhythms in performance." IN J.Aschoff (Ed). Handbook of Behavioural Neurobiology, Vol.4, New York: Plenum.

Corbera.X., Grau.C. (1993). "Diurnal type and hemispheric asymmetry," Cortex, 29(3), 519-528.

Coyle.K. (1989). "Circadian variation in cognitive functioning," PhD Thesis University of Wales Institute of Science and Technology.

Craig.A., Baer.K., Diekmann.A. (1981). "The effects of lunch on sensory-perceptual functioning in man," International Archives of Occupational and Environmental Health, 49, 105-114.

Craig.A., Condon.R. (1984). "Operational efficiency and time of day," Human Factors, 26, 197-205.

Craig.A., Condon.R. (1985). "Speed-accuracy trade-off and time of day," Acta Psychologia, 58, 115-122.

Craik.K. (1943). The nature of explanation. Cambridge: Cambridge University Press.

Curry.M.B., McDougall.S.J.P., de Bruijn.O. (1998). "The effects of the visual metaphor in determining icon efficacy," Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 1590 – 1594.

Damos.D. (1985). "The relation between the type A behaviour pattern, pacing and subjective workload under single- and dual-task conditions," Human Factors, 27(6), 675-680.

Davies.D.R. (Ed). (1984). Varieties of Attention. London: Academic Press.

Davies.D.R., Parasuraman.R., Toh.M.Y. (1984). "Time of day, memory load and vigilance performance." IN A.Mital (Ed). Trends in Ergonomics/Human Factors, Vol.1. Amsterdam: Elsevier.

DeGroot.A.D. (1965). Thought and choice in chess. Paris:Mouton.

Dijk.D.J., Duffy.J.F., Czeisler.C.A. (1992). "Ciracdian and sleep/wake dependent aspects of subjective alertness and cognitive performance," Journal of Sleep Research, 1, 112-117.

DiSessa.A.A. (1983). "A principled design for an integrated computational environment," Human-Computer Interaction, 1, 1-47.

Dinges.D., Orne.E., Evans.F., Orne.M. (1981). "Performance after naps in sleep-conducive and alerting environments." IN L. Johnson., D.Tepas., W.P. Colquhoun & M. Collogan (Eds). The Twenty-Four Hour Workday: Proceedings of a Symposium on Variations in Work-Sleep Schedules. Cincinnati: US Dept Health and Human Services.

Dinges.D., Orne.M., Orne.E. (1985). "Assessing performance upon abrupt awakening from naps during quasi-continuous operations," Behaviour Research Methods, 17(1), 37-45.

Dinges.D.F., Pack.F., Williams.K., Gillen.K.A., Powell.J.W., Ott.G.E., Aptowicz.C., Pack.A.I. (1997). "Cumulative sleepiness, mood disturbance and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night," Sleep, 20 (4), 267-277.

Duncan.J. (1984). "Selective attention and the organisation of visual information," Journal of Experimental Psychology: General, 113, 501-517.

Duncan.J., Humphreys.G.W. (1989). "Visual search and stimulus similarity," Psychological Review, 96, 433-458.

Duncan.J., Humphreys.G.W. (1992). "Visual search and attentional engagement," Journal of Experimental Psychology: Human Perception and Performance, 18, 578-588.

Easterbrook.J.A. (1959). "The effect of emotion on cue utilisation and the organisation of behaviour," Psychological Review, 66, 183-201.

Easterby.R.S. (1970). "The perception of symbols for machine displays," Ergonomics, 13(1), 149-158.

Ebbinghaus.H. (1885). "Memory: a contribution to experimental psychology." Republished in translation. New York: Dover Publications 1964.

Englund.C.E. (1979). "The diurnal function of reading rate, comprehension and efficiency," Report prepared for the Naval Health Research Centre, San Diego, California.

Eysenck.M.W., Folkard.S. (1980). "Personality, time of day and caffeine: some theoretical and conceptual problems in Revelle et al," Journal of Experimental Psychology: General, 109, 32-41.

Ferrara.M., De Gennaro.L., Bertini.M. (2000). "Time-course of sleep inertia upon awakening from nighttime sleep with different homeostatic conditions," Aviation, Space, Environmental Medicine, 71(3), 225-229.

Fisher.D.L., Tanner.N.S. (1992). "Optimal symbol set selection: a semi-automated procedure," Human Factors, 34, 79 – 95.

Folkard.S. (1975). "Diurnal variation in logical reasoning," British Journal of Psychology, 66, 1-8.

Folkard.S., Knauth.P., Monk.T.H., Rutenfranz.J. (1976). "The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system," Ergonomics, 19 (4), 479-488.

Folkard.S., Monk.T.H., Bradbury.R., Rosenthall.J. (1977). "Time of day effects in school children's immediate and delayed recall of meaningful material," British Journal of Psychology, 68, 45-50.

Folkard.S., Monk.T.H. (1978). "Time of day effects on immediate and delayed memory." IN M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds). Practical Aspects of Memory. London: Academic Press.

Folkard.S. (1979). "Time of day and level of processing," Memory and Cognition, 7, 247-252.

Folkard.S. (1980). "A note on 'Time of day effects in school childrens' immediate and delayed recall of meaningful material' – the influence of the importance of the information tested," British Journal of Psychology, 71, 95-97.

Folkard.S., Monk.T.H. (1980). "Circadian rhythms in human memory," British Journal of Psychology, 71, 295-307.

Folkard.S. (1981). "Circadian rhythms and shiftwork." IN G.R.J. Hockey (Ed). Stress and fatigue in human performance. New York: John Wiley.

Folkard.S., Monk.T.H. (1981). "Circadian rhythms in performance – one or more oscillators?" IN Sinz.R., Rosenwig.M.R. (Eds). Psychophysiology. Memory, motivation and event related potentials in mental operations. North Holland: Elsevier.

Folkard.S. (1982). "Circadian rhythms and human memory," IN F.M. Brown., R.C. Graeber. (Eds). Rhythmic Aspects of Behaviour. New Jersey: Lawrence Erlbaum Associates.

Folkard.S. (1983). "Diurnal variation." IN G.R.J. Hockey (Ed). Stress and Fatigue in Human Performance. Chichester: Wiley.

Folkard.S., Monk.T.H. (1983). "Chronopsychology: circadian rhythms and human performance. IN A. Gale & J.A. Edwards (Eds). Physiological correlates of human behaviour Vol. 2, Attention and Performance. London: Academic Press.

Folkard.S., Wever.R.A., Wildgraeber.C.M. (1983). "Multioscillatory control of circadian rhythms in human performance," Nature, 305, 223-226.

Folkard.S., Monk.T.H. (1984). "Do we need a multi-oscillator model of circadian rhythms in human performance?" IN E. Haus & H.F. Kabat (Eds.). Chronobiology. Karger: Basel.

Folkard.S., Marks.M., Minors.D., Waterhouse.J. (1985). "Circadian rhythms in human performance and affective state," Acta Psychiatry, 85, 568-581.

Folkard.S., Monk.T.H. (1985). "Circadian performance rhythms." IN S.Folkard & T.H.Monk (Eds). Hours of Work: Temporal Factors in Work Scheduling. Chichester: Wiley.

Folkard.S., Monk.T.H. (1987). "The measurement of circadian rhythms in psychological functioning." IN Schwing.L.E., Halberg.F., Ehret.C.F. (Eds). Chronobiotechnology and chronobiological engineering. Dordrecht, The Netherlands: Martinus Nijhoff, pp 189-201.

Folkard.S. Akerstedt.T. (1992). "A three-process model of the regulation of alertness-sleepiness." IN Broughton.R.J., Ogilvie.R.D. (Eds). Sleep, arousal and performance: a tribute to Bob Wilkinson. Boston: Birkhauser pp 11-26.

Folkard.S. (1996). "Body rhythms and shiftwork." IN Warr.P.B. (Ed). Psychology at Work (4th Ed), pp 39-72, Harmondsworth, Penguin Books.

Folkard.S. (1997). "Black times: temporal determinants of transport safety," Accident Analysis and Prevention, 29 (4), 417-430.

Folkard.S., Hill.J. (2002). "Shiftwork: body rhythm and social factors." IN Warr.P.B. (Ed). Psychology at Work (5th Ed), pp 51-76. London: Penguin Books.

Gander.P.H., Gregory.K.B., Graeber.R.C., Connell.L.J., Miller.D.L., Rosekind.M.R. (1998). "Flight crew IV: Overnight cargo operations," Aviation, Space and Environmental Medicine, 69 (9 sect 2), B26-B36.

Gander.P.H., Gregory.K.B., Graeber.R.C., Connell.L.J., Miller.D.L., Rosekind.M.R. (1998). "Flight crew fatigue V: Long-haul air transport operations," Aviation, Space and Environmental Medicine, 69 (9 sect 2), B37-B48.

Garcia.M., Badre.A.N., Stasko.J.T. (1994). "Development and Validation of Icons Varying in their Abstractness," Interacting with Computers, 6 (2), 191-211.

Gates.A.I. (1916). "Variations in efficiency during the day together with practice effects, sex differences and correlations," University of California Publications Psychology, 2, 1-156.

Geiselman.R.E., Landee.B.M., Christen.F.G. (1982). "Perceptual discriminability as a basis for selecting graphic symbols," Human Factors, 329 – 337.

Gilhooly.K.J., Logie.R.H. (1980). "Age-of-acquisition, imagery, concreteness, familiarity and ambiguity measures for 1,944 words," Behaviour Research Methods and Instrumentation, 12, 395-427.

Goldsmith.M. (1998). "What's in a location? Comparing object-based and space-based models of feature integration in visual search," Journal of Experimental Psychology: General, 127(2), 189-219.

Goonetilleke.R.S., Shih.H.M., On. H.K., Fritsch.J. (2001). "Effects of training and representational characteristics in icon design," International Journal of Human Computer Studies, 55 (5), 741-760.

Green.A.J.K., Barnard.P.J. (1990). "Iconic Interfacing: The Role of Icon Distinctiveness and Fixed or Variable Screen Locations," Human-Computer Interaction, Interact '90, 457-462.

Guerness.A., Leconte.L.C., Leconte.P. (1993). "Time of day effects on attention and memory efficiency: is chronopsychology a method for studying the functioning of the human subject?" Psychologia Beligica, 33 (2), 159-170.

Gunter.B., Furnham.A., Jarrett.J. (1984). "Personality, time-of-day and delayed memory for TV news," Journal of Personality and Individual Differences, 5 (1), 35-39.

Hakiel.S.R., Easterby.R.S. (1987). "Methods for the design and evaluation of icons for human-computer interfaces." IEEE 2nd International Conference on Command, Control, Communications and Management Information Systems. Bournemouth. UK.

Hancock.P.A. (1989). "A dynamic model of stress and sustained attention," Human Factors, 31(5), 519-537.

Hartley.L.R., Shirley.E. (1976). "Color-name interference at different times of day," Journal of Applied Psychology, 61, 119-122.

Hasher.L., Zacks.R.T., May.C.P. (1999). "Inhibitory control, circadian arousal and age," IN Gopher.D., Koriat.A. (Eds). Attention and performance XVII: cognitive regulation of performance: Interaction of theory and application. Attention and Performance, pp.653-675, Cambridge MA US: The MIT Press.

Heck.M.P. (1996). "Iconic verses textual representation within complex visual environments: which is the more effective communicator?" Dissertation Abstracts International: Section B: The Sciences and Enginnering, 57 (1-B), 0734.

Hemenway.K. (1982). "Psychological issues in the use of icons in command menus". Proceedings on Human Factors in Computer Systems. Gaithersburg, MD: ACM.

Hindmarch.I., Quinlan.P.T., Moore.K.L., Parkin.C. (1998). "The effects of black tea and other beverages on aspects of cognition and psychomotor performance," Psychopharmacology, 139(3), 230-238.

Hockey.G.R.J., Colquhoun.W.P. (1972). "Diurnal variation in human performance: a review." IN W.P. Colquhoun (Ed). Aspects of Human Efficiency: Diurnal Rhythm and Loss of Sleep. London: English Universities Press.

Hockey.G.R.J. (1984). "Varieties of attentional state: the effects of environment." IN R. Parasuraman & D.R. Davies (Eds). Varieties of Attention. London: Academic Press.

Hollingworth.H.L. (1914). "Variations in efficiency during the working day," Psychological Review, 21, 473-91.

Holloway.P.J. (1967). "The effect of lecture time on learning," British Journal of Educational Psychology, 36, 255-258.

Horne.J.A., Ostberg.O. (1976). "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms," International Journal of Chronobiology, 4, 97-110.

Horne.J.A., Brass.C.G., Pettit.A.N. (1980). "Circadian performance differences between morning and evening 'types'," Ergonomics, 23, 129-136.

Horne.J.A., Baumber.C.J. (1991). "Time of day effects of alcohol intake on simulated driving performance in women," Ergonomics, 34(11), 1377-1383.

Horne.J.A., Reyner.L.A. (1996). "Counteracting driver sleepiness: effects of napping, caffeine and placebo," Psychophysiology, 33 (3), 306-309.

Horton.W. (1994). The Icon Book: visual symbols for computer systems and documentation, Canada, John Wiley & Sons.

Hughes.D.G., Folkard.S. (1976). "Adaptation to an 8-hr shift in living routine by members of a socially isolated community," Nature, 264, 232-234.

Hutchins.E.L., Hollan.J.D., Norman.D.A., (1986). "Direct Manipulation Interfaces," IN Norman.D.A., Draper.S.W. (Eds), User Centered System Design. Hillsdale. NJ. Erlbaum.87-124.

Inouye.S.T., Kawamura.H. (1979). "Persistence of circadian rhythmicity in a mammalian hypothalamic 'island' containing the suprachiasmatic nuclei," Proceedings of the Natural Academy of Sciences USA, 76, 5962-5966.

Jaencke.L., Musial.F., Vogt.J., Kalveram.K.T. (1994). "Monitoring radio programs and time of day affect simulated car-driving performance," Perceptual and Motor Skills, 79 (1&2), 484-486.

Johnson.M.P., Duffy.J.F., Dijk.D.J., Ronda.J.M., Dyal.C.M., Czeisler.C.A. (1992). "Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature," Journal of Sleep Research, 1, 24-29.

Jones.S. (1983). "Stereotypy in pictograms of abstract concepts," Ergonomics, 26 (6), 605-611.

Kacmar.C.J., Carey.J.M. (1991). "Assessing the usability of icons in user interfaces". Behaviour and Information Technology, 10(6), 443-457.

Klein.K.E., Wegman.H.M., Hunt.B.I. (1972). "Desynchronisation as a function of body temperature and performance circadian rhythm as a result of outgoing and homecoming transmeridian flights," Aerospace Medicine, 43, 119-132.

Klein.K.E., Herrmann.R., Kuklinski.P., Wegmann.H.M. (1977). "Ciradian performance rhythms: experimental studies in air operations." IN R.R. Mackie (Ed).

Vigilance: theory, operational performance and physiological correlates. New York: Plenum Press.

Kleitman.N. (1939). "Sleep and wakefulness." Chicago: University of Chicago Press.

Kleitman.N., Jackson.D.P. (1950). "Body temperature and performance under different routines," Journal of Applied Psychology, 3, 309-328.

Kleitman.N. (1963). "Sleep and wakefulness. Revised and enlarged." Chicago: University of Chicago Press.

Kline.T.J., Ghali.L.M., Kline.D.W., Brown.S. (1990). "Visibility distance of highway signs among young, middle ages and older observers: Icons are better than text," Human Factors, 32(5), 609-619.

Lack.L.C., Lushington.K. (1996). "The rhythms of human sleep propensity and core body temperature," Journal of Sleep Research, 5, 1-11.

Laird.D.A. (1925). "Relative performance of college students as conditioned by time of day and day of week," Journal of Experimental Psychology, 8, 50-63.

Lamberti.D.M., Newsome.S.I. (1989). "Presenting abstract verses concrete information in expert systems: what is the impact on user performance". International Journal of Man-Machine Studies, 31, 27-45.

Lane.J.D. (1997). "Effects of brief caffeinated-beverage deprivation on mood, symptoms and psychomotor performance," Pharmacology, Biochemistry and Behaviour, 58 (1), 203-208.

Larkin.J.H. (1983). "The role of problem representation in physics". IN Gentner.D., Stevens.A.L. (Eds), Mental Models. Hillsdale, New Jersey: Erlbaum.

Lavie.P., Segal.S. (1989). "Twenty-four-hour structure of sleepiness in morning and evening persons investigated by ultrashort sleep-wake cycle," Sleep, 12 (6), 522-528.

Lawrence.J.B., Stanford.M.S. (1999). "Impulsivity and time of day: effects on performance and cognitive tempo," Personality and Individual Differences, 26(2), 199-207.

Lenne.M.G., Triggs.T.J., Redman.J.R. (1997). "Time of day variations in driving performance," Accident Analysis and Prevention, 29(4), 431-437.

Lenne.M.G., Triggs.T.J., Redman.J.R. (1998). "Interactive effects of sleep deprivation, time of day and driving experience on a driving task," Sleep, 21(1), 38-44.

Lewis.C. (1981). "Skill in algebra." IN Anderson.J.R. (Ed.), Cognitive Skills and their Acquisition. Hillsdale, New Jersey: Erlbaum.

Lieberman.H.R. (1992). "Caffeine." IN A.P. Smith., D.M. Jones (Eds). Handbook of Human Performance, Vol. 2, Health and Performance. London: Academic Press.

Linde.L. (1995). "Mental effects of caffeine in fatigued and non-fatigued female and male subjects," Ergonomics, 38(5), 864-885.

Lodding.K.N. (1983). "Iconic interfacing," IEEE Computer Graphics and Applications, 24, 11-20.

Lorenzetti.R., Natale.V. (1996). "Time of day and processing strategies in narrative comprehension," British Journal of Psychology, 87, 209-221.

Marks.M., Folkard.S. (1988). "The effects of time of day on recall from expository text." IN M.Gruneberg., P.Morris. & R.Sykes (Eds). Practical Aspects of Memory, Current Research and Issues, Vol.2. Chichester: Wiley.

Magyar.R.L. (1990). "Assessing icon appropriateness and icon discriminability with a paired-comparison testing procedure," Proceedings of the Human Factors Society 34th Annual Meeting, 1204 – 1208. Santa Monica, CA: Human Factors Society.

Manktelow.K., Jones.J. (1987). "Principles from the psychology of thinking and mental models". IN Gardiner.M.M., Christie.B. (Eds). Applying Cognitive Psycholgy to User-Interface Design. Chichester: Wiley and Sons. 83-177.

May.C.P., Hasher.L., Stoltzfus.E.R. (1993). "Optimal time of day and the magnitude of age differences in memory," Psychological Science, 4 (5), 326-330.

McDonald.N. (1989). "Fatigue and driving," Alcohol, Drugs and Driving, 5(3), 185-192.

McDougall.S.J.P., Curry.M.B., de Bruijn.O. (1996). "Effective symbology project: A review of symbol characteristics and their effects on usability," Unpublished report for BAE Systems.

McDougall.S.J.P., Curry.M.B., de Bruijn.O. (1999a). "Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity and semantic distance for 239 symbols", Behaviour Research Methods: Intsruments and Computers, 31(3), 487-519.

McDougall.S.J.P., de Bruijn.O. (1999b). "Effective symbology project: symbol usability and design advice and guidelines", Unpublished report for BAE Systems.

McDougall.S.J.P., Curry.M.B., de Bruijn.O. (2000). "Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity and distinctiveness," Journal of Experimental Psychology: Applied, 6 (4), 291-306.

McDougall.S.J.P., Curry.M.B., de Bruijn.O. (2001). "The effects of visual information on users' mental models: An evaluation of pathfinder analysis as a measure of icon usability," International Journal of Cognitive Ergonomics, 5 (1), 59-84.

McFadden.E., Tepas.D.I. (1997). "Effects of time of day and task demand on simulated sign recognition," Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, 1392 – 1396. Santa Monica, CA: Human Factors Society.

McLean.P.D. (1969). "Induced arousal and time of recall as determinants of paired-associate recall," British Journal of Psychology, 60, 57-62.

Mellette.H.C., Hutt.B.L., Askoritz.S.I., Horvath.S.M. (1951). "Diurnal variations in body temperature," Journal of Applied Physiology, 3, 665-75.

Merikle.P.M. (1980). "Selection from visual persistence by perceptual groups and category membership," Journal of Experimental Psychology: General, 109, 279-295.

Millar.K. (1979). "Word recognition in loud noise," Acta Psychologia, 43, 225-237.

Miller.K., Styles.B.C., Wastell.D.G. (1980). "Time of day and retrieval from long-term memory," British Journal of Psychology, 71, 407-414.

Miller.L.S., Lombardo.T.W., Fowler.S.C. (1995). "Caffeine and time of day effects on a force discrimination task in humans," Physiology and Behaviour, 57(6), 1117-1125.

Miller.L.A. (2001). "Predicting success in WIMP environments using isoperformance methodology," Dissertation Abstracts International: Section B: The Sciences and Enginnering, 61 (7-B), 3888.

Mills.T. (1967). "Man-machine communication and problem solving". IN Annual Review of Information Science and Technology, 2. NY: Interscience Publishers.

Mitchell.P.J., Redman.J.R. (1992). "Effects of caffeine, time of day and user history on study-related performance," Psychopharmacology, 109 (1-2), 121-126.

Monk.T.H. (1978). "Memory based performance measures in studies of shiftwork," Ergonomics, 21 (10), 819-826.

Monk.T.H. (1979). "Temporal effects in visual search." IN J.N. Clare., M.A. Sinclair (Eds). Search and the Human Observer. London: Taylor and Francis.

Monk.T.H. (1982). "The arousal model of time of day effects in human performance efficiency," Chronobiologia, 9 (1), 49-54.

Monk.T.H., Leng.V.C. (1982). "Time of day effects in simple repetitive tasks: some possible mechanisms," Acta Psychologia, 51, 207-221.

Monk.T.H., Weitzman.E., Fookson.J., Moline.M., Kronauer.R., Gander.P. (1983). "Task variables determine which biological clock controls circadian rhythms in human performance," Nature, 304, 543-545.

Monk.T.H., Fookson.J.E., Moline.M.L., Pollak.C.P. (1985). "Diurnal variation in mood and performance in a time isolated environment," Chronobiology International, 2, 185-193.

Monk.T.H., Leng.V.C.(1986). " Interactions between inter-individual and inter-task differences in the diurnal variation of human performance," Chronobiology International, 3, 171-177,

Monk.T.H. (1990). "The relationship of chronobiology to sleep schedules and performance demands," Work and Stress, 4(3), 227-236.

Monk.T.H., Buysse.D.J., Reynolds.C.F., et al (1996). "Circadian determinants of the post-lunch dip in performance," Chronobiology International, 13, 135-145.

Monk.T.H., Buysse.D.J., Reynolds III.C.F., Berga.S.L., Jarrett.D.B., Begley.A.E., Kupfer.D.J. (1997). "Circadian rhythms in human performance and mood under constant conditions," Journal of Sleep Research, 6, 9-18.

Moore.R.Y., Eichler.V.B. (1972). "Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat," Brain Research, 42, 201-206.

Moyes.J., Jordan.P.W. (1993). "Icon design and its effect on guessability, learnability and experienced user performance". IN Alty.J.L., Diaper.D., Guest.S. (Eds), People and Computers VIII. Cambridge, Cambridge University Press.

Muter.P., Mayson.C. (1986). "The role of graphics in item selection for menus," Behaviour and Information Technology, 5, 89-95.

Naitoh.P., Angus.R. (1989). "Napping and human functioning during prolonged work." IN D.Dinges & R.Broughton (Eds). Sleep and Alertness: Chronobiological, Behavioural and Medical Aspects of Napping. New York: Raven.

Nasanen.R., Ojanpaa.H., Kojo.I. (2001). "Effect of stimulus contrast on performance and eye movement in visual search," Vision Research, 41 (14), 1817-1824.

Natale.V., Lorenzetti.R. (1997). "Influences of morningness-eveningness and time of day on narrative comprehension," Personality and Individual Differences, 23 (4), 685-690.

Neisser.U. (1967). Cognitive Psychology. New York: Appleton-Century-Crofts.

Norman.D.A., Shallice.T. (1980). "Attention to action: willed and automatic control of behaviour," Centre for Human Information Processing. Technical Report No. 99

Oakhill.J. (1986a). "Effect of time of day and information importance on adult's memory for a short story," Quarterly Journal of Experimental Psychology, 38a, 419-430.

Oakhill.J. (1986b). "Effects of time of day on the integration of information in text," British Journal of Psychology, 77, 481-488.

Oakhill.J. (1987). "Interpreting elliptical verb phrases at different times of day: effects of plausibility and antecedent distance," Language and Speech, 30, 145-157.

Oakhill.J. (1988). "Effects of time on text memory and inference." IN M.M. Gruneberg., P.E. Morris & R.N. Sykes (Eds). Practical Aspects of Memory: current research and issues. Vol 2.

Oately.K., Goodwin.B.C. (1971). "The explanation and investigation of biological rhythms." IN W.P.Colquhoun (Ed). Biological Rhythms and Human Performance. London: Academic Press.

Owens.D.S., MacDonald.I., Tucker.P., Sytnik.N., Minors.D., Waterhouse.J., Totterdell.P., Folkard.S. (1998). "Diurnal trends in mood and performance do not all parallel alertness," Scandinavian Journal of Work, Environment and Health, 24 (3), 109-114.

Owens.D.S., Macdonald.I., Tucker.P., Sytnik.N., Totterdell.P., Minors.D., Waterhouse.D., Folkard.S. (2000). "Diurnal variations in the mood and performance of highly practised young women living under strictly controlled conditions," British Journal of Psychology, 91, 41-60.

Petros.T.V., Beckwith.B.E., Anderson.M. (1990). "Individual differences in the effects of time of day and passage difficulty on prose memory in adults," British Journal of Psychology, 81, 63-72.

Paivio.A., Yuille.J.C., Madiagn.S.A. (1968). "Concreteness, imagery and meaningfulness values for 925 nouns," Journal of Experimental Psychology: Monograph Supplement, 76, 1-25.

Paivio.A. (1971). "Imagery and verbal processes." Holt, Reinhart and Winston Inc. NY: USA.

Pittendrigh.C.S. (1981). "Circadian systems: general perspective." IN J. Aschoff (Ed). Handbook of Behavioural Neurobiology, Vol.4. Biological Rhythms. New York: Plenum Press.

Poirel.C., Larouche.B. (1987). "Human circadian patterns of memory processes: chronopsychology of rates of forgetting, a rhythmometric study," Psychological Reports, 61, 3-12.

Reason.J., Mycielska.K. (1982). "Absent-minded? The psychology of mental lapses and everyday errors." Englewood Cliffs, New Jersey: Prentice-Hall.

Revelle.W., Humphreys.M.S., Simon.L., Gilliland.K. (1980). "The interactive effect of personality, time of day and caffeine: a test of the arousal model," Journal of Experimental Psychology: General, 109, 1-31.

Rogers.Y. (1986). "Evaluating the meaningfulness of icon sets to represent command operations". IN Harrison.M.D., Monk.A,F, (Eds.), People and Computers: Designing for Usability. Cambridge: Cambridge University Press. 586-603.

Rogers.Y., Oborne.D.J.(1987). "Pictorial communication of abstract verbs in relation to human-computer interaction," British Journal of Psychology, 78, 99-112.

Rogers.Y. (1988). "Pictorial representations of abstract concepts in relation to human-computer interaction," PhD Thesis University of Wales.

Rogers. Y. (1989). "Icon design for the user interface". International Reviews of Ergonomics, 2, 129-154.

Rogers.Y. (1989). "Icons at the interface: their usefulness," Interacting with Computers, 1 (1), 105-117.

Rohr.G., Keppel.E. (1984). "Iconic Interfaces: Where to Use and How to Construct?" IN Hendrick.H.W., Brown.O.jr. (Eds.), Human Factors in Organisational Design and Management, North-Holland: Elsevier Science Publishers, 269-275.

Ryan.L., Hatfield.C., Hofstetter.M. (2002). "Caffeine reduces time-of-day effects on memory performance in older adults," Psychological Science, 13, 68-71.

Schmidt.S.R. (1996). "Category typicality effects in episodic memory: testing models of distinctiveness". Memory and Cognition, 24, 595 – 607.

Schwing.R.C. (1990). "Exposure-controlled highway fatality rates: temporal patterns compared to some explanatory variables," Alcohol, Drugs and Driving, 5(4) & 6(1), 275-285.

Sears.A., Jacko.J.A., Brewer.B., Robelo.L.D. (1998). "An empirical perspective on visual icon design," Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 448-452.

Sekuler.R., Blake.R. (1994). "Perception," (3rd Ed). London: McGraw-Hill.

Siffre.M. (1964). "Beyond time." New York: McGraw Hill.

Smilowitz.E.D. (1996). "Metaphors in user interface design: An empirical investigation," Dissertation Abstracts International: Section B: The Sciences and Enginnering, 57 (1-B), 0739.

Smith.D.C., Irby.C., Kimball.R., Verplank.B. (1982). "Designing the STAR interface," Byte, 7, 242-282.

Smith.A.P., Miles.C. (1985). "The combined effects of noise and nightwork on human function." IN D.Oborne (Ed). Contempory Ergonomics. London: Taylor & Francis.

Smith.A.P., Miles.C. (1986a). "The combined effects of nightwork and noise on human function." IN M.Haider., M.Koller & R.Cervinka (Eds). Studies in Industrial and Organisational Psychology 3. Night and Shiftwork: Long-term Effects and their prevention. Frankfurt: Peter Lang.

Smith.A.P., Miles.C. (1986b). "Effects of lunch on cognitive vigilance tasks," Ergonomics, 29, 1251-1261.

Smith.A.P. (1987a). "Activation states and semantic processing: a comparison of the effects of noise and time of day," Acta Psychologia, 54, 271-288.

Smith.A.P. (1987b). "Task-related motivation and time of testing." IN R.Schwarzer., H. van der Ploeg & C. Spielberger. (Eds). Advances in Test Anxiety Research, Vol.5. Lisse: Swets & Zeitlinger.

Smith.A.P., Miles.C. (1987a). "The combined effects of occupational health hazards: an experimental investigation of the effects of noise, nightwork and meals," International Archives of Occupational and Environmental Health, 59, 83-89.

Smith.A.P., Miles.C. (1987b). "Effects of lunch on selective and sustained attention," Neuropsychobiology, 16, 117-120.

Smith.A.P. (1988). "Effects of meals on memory and attention." IN M.M. Gruneberg., P.Morris & R.N.Sykes (Eds). Practical Aspects of Memory: Current Research and Issues, Vol.2. Chichester: Wiley.

Smith.A.P. (1989). "Circadian variation in stress: the effects of noise at different times of the day." IN C.D. Spielberger., J. Strelau (Eds). Stress and Anxiety, Vol. 12. New York: McGraw-Hill.

Smith.A.P., Rusted.J.M., Eaton-Williams.P., Savory.M., Leathwood.P. (1990). "Effects of caffeine given before and after lunch on sustained attention," Neuropsychobiology, 23 (3), 160-163.

Smith.A.P. (1991). "Strategy choice and time of day." IN E.J. Lovesey. Contemporary Ergonomics. London: Taylor & Francis.

Smith.A.P., Rusted.J.M., Savory.M., Eaton-Williams.P., Hall.S.R. (1991). "The effects of caffeine, impulsivity and time-of-day on performance, mood and cardiovascular function," Journal of Psychopharmacology, 5(2), 120-128.

Smith.A.P. (1992). "The effects of time of day, introversion and neuroticism on selectivity in memory and attention," Perceptual and Motor Skills, 74, 851-860.

Smith.A.P., Kendrick.A. (1992). "Meals and performance." IN D.M. Jones & A.P. Smith (Eds). A Handbook of Human Performance, Vol.2. London: Academic Press.

Smith.A.P., Brockman.P., Flynn.R., Maben.A. (1993). "Investigation of the effects of coffee on alertness and performance during the day and night," Neuropsychobiology, 27 (4), 217-223.

Smith.A.P., Clark.R., Gallagher.J. (1999). "Breakfast cereal and caffeinated coffee: effects on working memory, attention, mood and cardiovascular function," Physiology and Behaviour, 67(1), 9-17.

Smith.A.P. (2000). "Breakfast cereal, caffeinated coffee, mood and cognition," Nutrition, 16(3), 228-229.

Snodgrass.J.G., Vanderwart.M. (1980). "A standardised set of 260 pictures: norms for name agreement, familiarity and visual complexity," Journal of Experimental Psychology: Human Learning and Memory, 6, 147-215.

Stammers.R.B., George.D.A., Carey.M.S. (1989). "An Evaluation of Abstract and Concrete Icons for a CAD Package". IN E.D. Megaw (Ed). Contemporary Ergonomics 1989. London: Taylor and Francis.

Stammers.R.B. (1990). "Judged Appropriateness of Icons as a Predictor of Identification Performance." IN E.J. Lovesay (Ed). Contemporary Ergonomics. London: Taylor and Francis.

Stephen.F.K., Zucker.I. (1972). "Circadian rhythms in drinking behaviour and locomotory activity of rats are eliminated by hypothalamic lesions," Proceedings of the National Academy of Sciences USA, 69, 1583-1586.

Stotts.D.B. (1998). "The Usefulness of Icons on the Computer Interface: Effect of Graphical Abstraction and Functional Representation on Experienced and Novice Users," Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 453-457.

Streitz.N., Liesser.A., Wolters.A. (1989). "The combined effect of metaphor worlds and dialogue, odes in human-computer interaction. IN Klix.F., Streitz.N., Waern.Y., Wandke.H. (Eds.), Man-Computer Research MACINTER II, North-Holland, Amsterdam, 75-88.

Summala.H., Mikkola.T. (1994). "Fatal accidents among car and truck drivers: effects of fatigue, age and alcohol consumption," Human Factors, 36 (2), 315-326.

Testu.F., Clarisse.R. (1999). "Time-of-day and day-of-week effects on mnemonic performance," Chronobiology International, 16 (4), 491-503.

Thayer.R.E. (1967). "Measurement of activation through self-report," Psychological Reports, 20, 663-678.

Thayer.R.E. (1978). "Towards a psychological theory of multidimensional activation (arousal)," Motivation and Emotion, 2, 1-34.

Thayer.R.E. (1989). "The biopsychology of mood and arousal." New York: Oxford University Press.

Tilley.A., Warren.P. (1983). "Retrieval from semantic memory at different times of day," Journal of Experimental Psychology, 9, 718-724.

Treisman.A.M., Gelade.G. (1980). "A feature-integration theory of attention," Cognitive Psychology, 12, 97-136.

Treisman.A., Kahneman.D., Burkell.J. (1983). "Perceptual objects and the cost of filtering," Perception and Psychophysics, 33, 527-532.

Treisman.A. (1986). "Features and objects in visual processing," Scientific American, 255, 114-125.

Walker.E.L., Tarte.R.D. (1963). "Memory storage as a function of time with homogenous lists," Journal of Verbal Learning and Verbal Behaviour, 2, 113-119.

Waterhouse.J., Weinert.D., Minors.D., Folkard.S., Owens.D., Atkinson.G., Nevill.A., Reilly.T. (2000). "Estimates of the daily phase and amplitude of the endogenous component of the circadian rhythm of core temperature in sedentary humans living nychthermerally," Biological Rhythms Research, 31(1), 88-107.

Waters.G.S., Seidenberg.M.S., Bruck.M. (1984). "Children's and adult's use of spelling-sound information in three reading tasks". Memory and Cognition, 12, 293 – 305.

Webb.W., Agnew.H. (1978). "Effects of rapidly rotating shifts on sleep patterns and sleep structure," Aviation, Space and Environmental Medicine, 49(2), 384-389.

Wever.R.A.(1979). "The circadian system of man: results of experiments under temporal isolation." New York, Springer-Verlag.

Wever.R.A. (1982). "Behavioural aspects of circadian rhythmicity." IN F.M. Brown., R.C.Graeber. (Eds). Rhythmic Aspects of Behaviour. London: Lawrence Erlbaum & Associates.

White.N.M. (1998). "Cognitive enhancement: an everyday event?" International Journal of Psychology, 33(2), 95-105.

Whitney.P., Williams.D.L. "Circadian variations in semantic encoding functions," Unpublished Manuscript.

Wiedenbeck.S. (1999). "The use of icons and labels in an end user application program: An empirical study of learning and retention," Behaviour and Information Technology, 18 (2). 68-82.

Winch.W.H. (1912). "Mental fatigue in day school children as measured by immediate memory, part 1," Journal of Educational Psychology, 3, 18-28.

Winocur.G., Hasher.L. (1999). "Aging and time of day effects on cognition in rats," Behavioural Neuroscience, 113(5), 991-997.

Wood.W.T., Wood.S.K. (1987). "Icons in everyday life". IN G. Salvendy, S.L. Sauter & J.J. Hurrell. (Eds). Social Ergonomic and Stress Aspects of Work with Computers. Amsterdam, Elsevier, pp. 97-104.

Yerkes.R.M., Dodson.J.D. (1908). "The relation of strength of stimulus to rapidity of habit formation," Journal of Comparative and Neurological Psychology, 18, 459-482.

Yoon.C., May.C.P., Hasher.L. (2000). "Ageing, circadian arousal patterns, and cognition." IN D. Park., N. Schwarz (Eds). Cognitive Ageing: A Primer. Sussex: Psychology Press.

Zhou.W. (2000). "The role of cognition in the use and design of computer image icons," Dissertation Abstracts International: Section B: The Sciences and Enginnering, 61 (5-B), 2793.

Ultimate Resource Locators

www.mypharmacy.co.uk/health_products/products/b/braun/braun_thermoscan_plus.htm

Appendix 1:Experiment 1

(a):ANOVA Summary Tables for Experiment 1

Repeated Measures ANOVA Summary Table/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	824865.03	4	206216.26	192	3.39	0.011
TOD*Condition	207210.93	4	51802.73	192	0.851	0.495
Complexity	32826152.54	1	32826152.54	48	477.10	0.000
TOD*Complexity	239928.35	4	59982.09	192	3.76	0.006
Complexity*Condition	28757.80	1	28757.80	48	0.42	0.521
TOD*Complexity*Condition	40424.05	4	10106.01	192	0.63	0.640
Presence	30681554.59	1	30681554.59	48	480.51	0.000
TOD*Presence	19084.64	4	4771.16	192	0.69	0.602
Presence*Condition	15824.25	1	15824.245	48	0.25	0.621
TOD*Presence*Condition	83395.39	4	20848.85	192	3.00	0.020
Complexity*Presence	12519.64	1	12519.64	48	1.20	0.279
TOD*Complexity*Presence	62178.39	4	15544.60	192	2.05	0.090
Complexity*Presence*Condition	3844.78	1	3844.78	48	0.37	0.547
TOD*Complexity*Presence*Condition	17168.04	4	4292.01	192	0.57	0.688
B/S Factor(s)						
Condition	2028114.30	1	2028114.30	48	1.97	0.167

Repeated Measures ANOVA Summary Table/Time-of-Day/Temperatures

	SS	*df*	*MS*	*Error*	*F*	*P*
W/S Factor(s)						
TOD	18.83	4	4.71	192	27.87	0.000
Temperature Before/After	48.80	1	48.80	48	115.60	0.000
TOD*Temperature Before/After	0.81	4	0.20	192	1.92	0.109
Temperature Before/After*Condition	0.24	1	0.24	48	0.57	0.453
TOD*Temperature Before/After*Condition	0.53	4	0.13	192	1.26	0.287
B/S Factor(s)						
Condition	0.74	1	0.74	48	0.36	0.553

(b): Simple Interaction Effects/Simple Simple Main Effects/Simple Main Effects Summary Tables for Experiment 1

Simple Main Effect Summary Table (Time-of-Day x Complexity)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD*Complex	913362.50	4	228340.62	192	4.19	0.003
TOD*Complex*Condition	110945.05	4	27736.26	192	0.51	0.729
TOD*Simple	151430.88	4	37857.72	192	1.69	0.153
TOD*Simple*Condition	136689.93	4	34172.48	192	1.53	0.195

Simple Interaction Effect Summary Table (Time-of-Day x Presence x Condition)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Unequal Condition *TOD	1319990.64	4	329997.66	192	2.71	0.031
Unequal Condition *Presence	32090953.29	1	32090953.29	48	251.29	0.000
Unequal Condition *TOD*Presence	159963.12	4	39990.78	192	2.88	0.024
Equal Condition *TOD	744161.28	4	186040.32	192	1.53	0.196
Equal Condition *Presence	29303804.37	1	29303804.37	48	229.47	0.000
Equal Condition *TOD*Presence	44996.92	4	11249.23	192	0.81	0.520

Simple Simple Main Effect Summary Table (Time-of-Day x Presence x Condition)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Unequal Condition *TOD*Present Icons	219004.89	4	54751.22	192	1.64	0.166
Unequal Condition *TOD*Absent Icons	520971.99	4	130243.00	192	3.78	0.006
Equal Condition *TOD*Present Icons	265624.35	4	66406.09	192	1.99	0.098
Equal Condition *TOD*Absent Icons	128954.75	4	32238.69	192	0.94	0.445

Simple Simple Main Effect Summary Table (Time-of-Day x Presence x Condition)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
0900*Present Icons	167295987.10	1	167295987.10	48	1820.57	0.000
0900*Present Icons*Condition	345815.74	1	345815.74	48	3.76	0.58
0900*Absent Icons	267938235.50	1	267938235.50	48	1496.51	0.000
0900*Absent Icons*Condition	375333.90	1	375333.90	48	2.10	0.154
1200*Present Icons	167044044.40	1	167044044.40	48	1869.47	0.000
1200*Present Icons*Condition	202239.08	1	202239.08	48	2.26	0.139
1200*Absent Icons	269949861.90	1	269949861.90	48	1744.08	0.000

1200*Absent Icons*Condition	356871.23	1	356871.23	48	2.31	0.135
1500*Present Icons	183404154.50	1	183404154.50	48	1477.37	0.000
1500*Present Icons*Condition	472209.48	1	472209.48	48	3.80	0.057
1500*Absent Icons	294093723.00	1	294093723.00	48	1544.36	0.000
1500*Absent Icons* Condition	105221.74	1	105221.74	48	0.55	0.461
1800*Present Icons	184015709.10	1	184015709.10	48	1696.23	0.000
1800*Present Icons*Condition	193078.51	1	193078.51	48	1.78	0.188
1800*Absent Icons	287129331.00	1	287129331.00	48	1541.95	0.000
1800*Absent Icons*Condition	196109.92	1	196109.92	48	1.05	0.310
2100*Present Icons	175641611.40	1	175641611.40	48	2051.75	0.000
2100*Present Icons*Condition	81953.38	1	81953.38	48	0.96	0.333
2100*Absent Icons	283682573.00	1	283682573.00	48	1803.81	0.000
2100*Absent Icons*Condition	5711.88	1	5711.88	48	0.04	0.850

(c): Mean Percentage Accuracy Summary Tables for Experiment 1

Mean Percentage Accuracy Rates at Each Time-of-Day for all Participants/Complex Trials

Condition (A= unequal vs B=equal)	*Participant No.*	*Mean % Accuracy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @1800*	*Mean % Accuracy @ 2100*	*Total Across all TOD for Each Participant*
A	1	97.22	95.56	95.56	98.33	96.11	96.56
A	3	87.78	83.89	90.56	87.78	90.00	88.00
A	5	87.22	90.56	92.78	90.00	92.22	90.56
A	7	95.56	93.89	93.89	97.22	98.89	95.89
A	9	96.11	97.22	95.56	97.22	99.44	97.11
A	11	91.11	91.11	90.00	93.33	87.78	90.67
A	13	93.33	94.44	93.89	89.44	89.44	92.11
A	15	93.89	90.56	86.67	83.89	88.33	88.67
A	17	91.11	88.89	91.11	95.56	92.78	91.89
A	19	97.22	97.78	91.11	95.00	95.00	95.22
A	21	98.33	98.33	97.22	97.78	97.22	97.78
A	23	89.44	79.44	83.89	85.56	80.56	83.78
A	25	98.89	97.78	98.33	97.78	96.67	97.89
A	27	96.67	98.33	98.89	98.33	99.44	98.33
A	29	96.67	95.00	96.67	98.33	95.56	96.45
A	31	89.44	88.33	87.78	89.44	93.33	89.66
A	33	95.00	92.78	97.22	98.89	93.33	95.44
A	35	97.78	94.44	95.00	92.22	96.11	95.11
A	37	98.33	97.78	97.22	97.78	98.89	98.00
A	39	97.22	99.44	97.78	99.44	98.33	98.44
A	41	92.22	86.67	85.00	92.22	90.00	89.22
A	43	91.11	87.22	87.78	87.78	83.89	87.56
A	45	94.44	93.89	95.00	91.67	96.11	94.22
A	47	94.44	91.67	95.56	95.56	92.78	94.00
A	49	94.44	93.33	93.33	97.22	93.89	94.44
Totals		94.20	92.73	93.11	93.91	93.44	
Grand total over all times of day & over all participants							93.91

B	2	84.44	96.11	87.78	83.89	88.89	88.22
B	4	98.89	97.78	93.33	95.56	97.22	96.56
B	6	96.11	99.44	98.89	98.33	99.44	98.44
B	8	89.44	95.00	93.33	93.89	91.11	92.55
B	10	96.67	95.00	92.22	97.22	92.78	94.78
B	12	97.78	98.89	95.56	95.00	96.11	96.67
B	14	95.00	94.44	93.33	93.89	91.67	93.67
B	16	98.33	92.78	93.89	92.78	97.78	95.11
B	18	97.22	96.11	96.67	89.44	95.56	95.00
B	20	95.00	93.33	82.22	90.00	91.67	90.44
B	22	97.78	97.78	98.89	97.78	97.78	98.00
B	24	92.78	98.89	96.11	93.33	96.67	95.56
B	26	90.56	92.78	95.56	90.00	91.11	92.00
B	28	90.56	93.33	91.67	93.89	92.78	92.45
B	30	95.00	96.67	90.56	95.56	91.67	93.89
B	32	95.56	95.56	96.11	96.67	96.67	96.11
B	34	92.22	91.11	86.11	92.22	91.67	90.67
B	36	96.67	98.89	97.22	97.78	97.78	97.67
B	38	96.11	97.78	98.89	98.33	98.33	97.89
B	40	95.56	96.11	97.22	95.00	93.33	95.44
B	42	98.33	97.22	98.33	98.89	96.67	97.89
B	44	95.56	93.89	93.33	96.11	94.44	94.67
B	46	98.33	100.00	99.44	97.22	95.00	98.00
B	48	90.00	89.44	95.00	92.22	86.67	90.67
B	50	96.67	94.44	97.22	97.22	93.89	95.89
Totals		94.82	95.71	94.36	94.49	94.27	
Grand total over all times of day & over all participants							94.73
Overall mean % accuracy over all times of day, over all participants & both conditions							94.10

Mean Percentage Accuracy Rates at Each Time-of-Day for all Participants/Simple Trials

Condition (A= unequal vs B=equal)	*Participant No.*	*Mean % Accuracy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @1800*	*Mean % Accuracy @ 2100*	*Total across all TOD for Each Participant*
A	1	97.78	95.56	95.56	95.00	95.00	95.78
A	3	86.67	81.11	91.11	86.11	86.67	86.33
A	5	91.11	94.44	95.56	92.22	93.33	93.33
A	7	98.89	95.56	96.67	97.78	97.22	97.22
A	9	95.56	97.22	95.56	96.11	97.78	96.45
A	11	91.67	92.78	92.78	92.22	91.67	92.22
A	13	94.44	93.89	96.11	92.78	89.44	93.33
A	15	97.22	96.67	96.67	95.56	94.44	96.11
A	17	92.78	91.67	93.33	93.33	89.44	92.11
A	19	98.33	98.33	95.00	98.89	97.78	97.67
A	21	98.33	97.22	97.22	95.00	97.22	97.00
A	23	88.89	87.22	86.67	85.56	80.56	85.78
A	25	98.33	98.33	98.89	98.33	97.78	98.33
A	27	99.44	98.89	98.89	98.33	100.00	99.11
A	29	98.33	96.67	96.67	98.89	96.11	97.33
A	31	94.44	93.33	91.11	92.22	92.78	92.78
A	33	98.89	96.67	97.78	96.67	96.67	97.34
A	35	96.11	97.78	96.67	100.00	99.44	98.00
A	37	99.44	99.44	98.33	98.33	98.89	98.89
A	39	97.22	100.00	100.00	98.89	100.00	99.22
A	41	90.56	92.78	89.44	89.44	91.67	90.78
A	43	92.22	89.44	89.44	93.89	95.00	92.00
A	45	98.89	99.44	99.44	97.78	95.56	98.22
A	47	93.89	91.11	93.89	95.56	100.00	94.89
A	49	98.89	98.33	98.33	97.22	96.67	97.89
Totals		95.53	94.96	95.24	95.04	94.84	
Grand total over all times of day & over all participants							95.12

B	2	91.67	91.67	93.33	91.67	95.00	92.67
B	4	96.67	93.33	93.33	100.00	95.00	95.67
B	6	100.00	96.67	100.00	100.00	100.00	99.33
B	8	90.00	95.00	88.33	93.33	95.00	92.33
B	10	98.33	98.33	96.67	100.00	98.33	98.33
B	12	100.00	96.67	98.33	100.00	96.67	98.33
B	14	100.00	93.33	95.00	93.33	95.00	95.33
B	16	98.33	95.00	98.33	98.33	100.00	98.00
B	18	100.00	100.00	96.67	95.00	98.33	98.00
B	20	96.67	96.67	80.00	88.33	93.33	91.00
B	22	100.00	100.00	100.00	100.00	100.00	100.00
B	24	96.67	95.00	96.67	95.00	98.33	96.33
B	26	95.00	98.33	96.67	98.33	98.33	97.33
B	28	88.33	93.33	88.33	91.67	93.33	91.00
B	30	98.33	98.33	96.67	96.67	100.00	98.00
B	32	93.33	96.67	91.67	90.00	96.67	93.67
B	34	95.00	96.67	100.00	93.33	93.33	95.67
B	36	100.00	100.00	98.33	100.00	100.00	99.67
B	38	98.33	98.33	96.67	100.00	98.33	98.33
B	40	100.00	93.33	93.33	96.67	91.67	95.00
B	42	100.00	98.33	98.33	98.33	100.00	99.00
B	44	95.00	86.67	88.33	95.00	91.67	91.33
B	46	100.00	100.00	100.00	98.33	96.67	99.00
B	48	95.00	98.33	96.67	95.00	91.67	95.33
B	50	96.67	98.33	96.67	100.00	98.33	98.00
Totals		96.93	96.33	95.13	96.33	96.60	

Grand total over all times of day & over all participants 96.27

Overall mean % accuracy over all times of day,
over all participants & both conditions 95.69

Appendix 2: Experiment 2

(a): ANOVA Summary Tables for Experiment 2

Repeated Measures ANOVA Summary Table/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	596725.86	3	198908.62	138	6.02	0.001
TOD*Condition	49135.51	3	16378.50	138	0.50	0.686
Presence	6801966.17	1	6801966.17	46	330.70	0.000
TOD*Presence	5330.34	3	1776.78	138	0.58	0.632
Presence*Condition	142820.42	1	142820.42	46	6.94	0.011
TOD*Presence*Condition	1304.50	3	434.83	138	0.14	0.935
B/S Factor(s)						
Condition	29293216.95	1	29293216.95	46	84.77	0.000

Repeated Measures ANOVA Summary Table/Time-of-day/Temperatures

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	38.91	3	12.97	138	116.24	0.000
TOD*Condition	0.77	3	0.26	138	2.29	0.081
Temperature Before/After	6.54	1	6.54	46	124.71	0.000
TOD*Temperature Before/After	0.24	3	0.08	138	4.06	0.008
Temperature Before/After*Condition	0.02	1	0.02	46	0.36	0.551
TOD*Temperature Before/After*Condition	0.08	3	0.03	138	1.38	0.251
B/S Factor(s)						
Condition	1.09	1	1.09	0.72	1.52	0.224

(b): Simple Interaction Effects/Simple Simple Main Effects/Simple Main Effects Summary Tables for Experiment 2

Simple Main Effects Summary Table (Presence x Condition)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Present Icons*Condition	13941933.01	1	13941933.01	46	67.73	0.000
Absent Icons*Condition	15368656.04	1	15368656.04	46	88.58	0.000

Simple Main Effects Summary Table (Time-of-Day*Temperature)/Temperature

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
0900*Temperature before/after	1746469.56	1	1746469.56	46	231.40	0.000
0900*Temperature before/after*Condition	26514.40	1	26514.40	46	3.51	0.067
1200*Temperature before/after	1814488.01	1	1814488.01	46	305.75	0.000
1200*Temperature before/after*Condition	38677.94	1	38677.94	46	6.52	0.014
1500*Temperature before/after	1557134.44	1	1557134.44	46	181.59	0.000
1500*Temperature before/after*Condition	45021.61	1	45021.61	46	5.25	0.027
1800*Temperature before/after	1689204.50	1	1689204.50	46	217.27	0.000
1800*Temperature before/after*Condition	33910.97	1	33910.97	46	4.36	0.042

(c): Mean Percentage Accuracy Summary Tables for Experiment 2

Mean Percentage Accuracy Rates at Each Time-of-Day for all Participants

Condition (A=multi-feature vs B=gestalt)	*Participant No.*	*Mean % Accuracy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @ 1800*	*Total Across all TOD for Each Participant*
A	1	87.78	95.00	89.44	85.56	89.45
A	3	96.67	95.56	96.11	98.89	96.81
A	5	96.67	97.78	95.56	96.11	96.53
A	7	93.89	92.78	96.67	92.22	93.89
A	9	98.33	97.22	95.56	98.33	97.36
A	11	95.00	95.56	91.67	92.78	93.75
A	13	96.67	95.56	97.22	97.22	96.67
A	15	92.78	96.11	95.56	94.44	94.72
A	17	100.00	92.78	98.33	96.67	96.95
A	19	97.22	96.11	94.44	95.56	95.83
A	21	91.11	90.56	93.89	95.56	92.78
A	23	95.56	92.78	96.11	93.33	94.45
A	25	88.33	86.67	86.11	89.44	87.64
A	27	98.33	96.11	95.56	92.22	95.56
A	29	98.89	97.22	97.22	97.22	97.64
A	31	93.33	84.44	87.22	85.56	87.64
A	33	96.11	88.89	90.56	92.22	91.95
A	35	79.44	77.78	85.00	86.67	82.22
A	37	93.89	89.44	92.78	87.78	90.97
A	39	99.44	97.78	99.44	99.44	99.03
A	41	95.00	93.89	96.11	91.67	94.67
A	43	95.56	93.89	97.78	93.89	95.28
A	45	91.67	95.56	97.22	95.56	95.00
A	47	96.67	97.78	97.22	96.67	97.09
Totals		94.51	93.22	94.28	93.54	
Grand total over all times of day & over all participants						93.91

B	2	93.89	95.56	95.00	95.56	95.00
B	4	90.56	93.89	91.67	89.44	91.39
B	6	92.78	90.56	89.44	86.67	89.86
B	8	98.33	97.22	91.67	94.44	95.42
B	10	95.56	93.33	97.22	93.33	94.86
B	12	97.78	97.22	97.78	96.67	97.36
B	14	98.33	99.44	98.33	99.44	98.89
B	16	97.78	100.00	97.78	95.00	97.64
B	18	99.44	96.67	95.56	96.67	97.09
B	20	98.33	98.89	95.00	94.44	96.67
B	22	97.22	96.67	95.56	96.11	96.39
B	24	96.67	94.44	93.89	93.33	94.58
B	26	96.67	93.89	96.11	95.56	95.56
B	28	95.00	93.33	93.89	95.56	94.45
B	30	93.89	96.11	92.78	97.22	95.00
B	32	96.67	97.78	98.89	95.00	97.09
B	34	98.89	97.78	98.33	97.78	98.20
B	36	95.00	94.44	94.44	92.78	94.17
B	38	98.33	98.33	97.78	99.44	98.47
B	40	94.44	95.56	93.89	93.33	94.31
B	42	99.44	97.22	97.78	96.67	97.78
B	44	96.67	93.89	95.56	96.67	95.70
B	46	93.33	91.11	90.56	91.11	91.53
B	48	96.11	97.22	96.11	97.78	96.81
Totals		96.30	95.86	95.21	95.00	
Grand total over all times of day & over all participants						95.59
Overall mean % accuracy over all times of day, over all participants & both conditions						94.74

Appendix 3: Experiments 3 & 4

(a): ANOVA Summary Tables for Experiments 3 & 4

Repeated Measures ANOVA Summary Table/Time-of-Day/ Response Times/Exp. 3

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	280557.64	3	93519.21	69	4.01	0.011
Concreteness	1731.64	1	1731.64	23	0.18	0.678
TOD*Concreteness	39141.44	3	13047.15	69	2.04	0.116
Complexity	856987.59	1	856987.59	23	121.97	0.000
TOD*Complexity	13526.87	3	4508.96	69	0.75	0.527
Block	41651.65	2	20825.83	46	3.02	0.059
TOD*Block	21976.06	6	3662.68	138	0.515	0.796
Concreteness*Complexity	30046.71	1	30046.71	23	5.04	0.035
TOD*Concreteness*Complexity	16320.11	3	5440.04	69	1.14	0.339
Concreteness*Block	19904.60	2	9952.30	46	1.45	0.245
TOD*Concreteness*Block	9752.05	6	1625.34	138	0.272	0.949
Complexity*Block	5659.63	2	2829.82	46	0.890	0.418
TOD*Complexity*Block	25358.09	6	4226.35	138	0.845	0.537
Concreteness*Complexity*Block	9123.19	2	4561.59	46	0.854	0.433
TOD*Concreteness*Complexity*Block	15969.37	6	2661.56	138	0.573	0.752

Repeated Measures ANOVA Summary Table/Time-of-Day/ Response Times/Exp. 4

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	557378.48	3	185792.83	69	1.14	0.339
Concreteness	1496046.36	1	1496046.36	23	14.14	0.001
TOD*Concreteness	165854.86	3	55284.96	69	1.99	0.124
Complexity	2324901.10	1	2324901.10	23	53.89	0.000
TOD*Complexity	78536.93	3	26178.98	69	1.51	0.220
Block	1068981.85	2	534490.93	46	24.80	0.000
TOD*Block	108100.37	6	18016.73	138	0.76	0.602
Concreteness*Complexity	221764.13	1	221764.13	23	6.45	0.018
TOD*Concreteness* Complexity	72664.58	3	24221.53	69	1.60	0.198
Concreteness*Block	167832.31	2	83916.16	46	4.48	0.017
TOD*Concreteness*Block	90308.000	6	15051.33	138	0.98	0.444
Complexity*Block	53726.76	2	26863.38	46	1.56	0.220
TOD*Complexity*Block	78522.13	6	13087.02	138	1.38	0.225
Concreteness*Complexity *Block	28033.25	2	14016.63	46	0.82	0.447
TOD*Concreteness* Complexity*Block	39304.99	6	6550.83	138	0.34	0.917

Repeated Measures ANOVA Summary Table/Time-of-Day/ Temperature/Exp. 3

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	18.32	3	6.12	69	42.79	0.000
Before/After	9.54	1	9.54	23	58.96	0.000
TOD*Before/After	0.22	3	0.07	69	1.60	0.197

Repeated Measures ANOVA Summary Table/Time-of-Day/ Temperature/Exp. 4

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	14.84	3	4.95	69	42.82	0.000
Before/After	13.71	1	13.71	23	74.85	0.000
TOD*Before/After	0.22	3	0.07	69	1.34	0.269

(b): Simple Interaction Effects/Simple Simple Main Effects/Simple Main Effects Summary Tables for Experiments 3 & 4

Simple Main Effects Summary Table (Concreteness x Complexity)/Time-of-Day/Response Times/Exp. 3

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Abstract Icons*Complexity	603984.14	1	603984.14	23	123.38	0.000
Concrete Icons*Complexity	283050.16	1	283050.16	23	34.96	0.000

Simple Main Effects Summary Table (Concreteness x Complexity)/Time-of-Day/Response Times/Exp.4

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Concreteness*Simple Icons	282910.96	1	282910.96	23	4.07	0.056
Concreteness*Complex Icons	1434899.52	1	1434899.52	23	20.31	0.000

Simple Main Effects Summary Table (Concreteness x Block)/Time-of-Day/Response Times/Exp. 4

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Abstract Icons*Block	1029604.91	2	514802.46	46	23.38	0.000
Concrete Icons*Block	207209.25	2	103604.63	46	5.67	0.006

Simple Main Effects Summary Table (Concreteness x Block)/Time-of-Day/Response Times/Exp. 4

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Concreteness*Block 1	1070448.19	1	1070448.19	23	14.48	0.001
Concreteness*Block 2	237284.23	1	237284.23	23	7.19	0.013
Concreteness*Block 3	356146.25	1	356146.25	23	9.80	0.005

(c): Mean Percentage Accuracy Summary Tables for Experiments 3 & 4

Mean Percentage Accuracy Rates at Each Time-of-Day for all Participants/ Exp. 3

Participant Number	*Mean % Accuracy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @ 1800*	*Total across all TOD for Each Participant*
1	100.00	98.61	99.54	99.54	99.42
2	100.00	99.07	100.00	99.07	99.54
3	100.00	99.07	99.07	100.00	99.54
4	100.00	99.54	99.54	100.00	99.77
5	99.54	99.07	100.00	99.07	99.42
6	100.00	99.54	98.61	99.54	99.42
7	100.00	100.00	100.00	100.00	100.00
8	100.00	100.00	99.54	99.54	99.77
9	100.00	100.00	100.00	100.00	100.00
10	100.00	100.00	99.54	100.00	99.88
11	100.00	100.00	100.00	100.00	100.00
12	100.00	99.54	100.00	100.00	99.88
13	100.00	100.00	100.00	100.00	100.00
14	100.00	100.00	100.00	100.00	100.00
15	99.07	98.15	98.15	93.98	97.34
16	100.00	100.00	100.00	99.54	99.88
17	100.00	100.00	100.00	100.00	100.00
18	99.54	99.07	100.00	99.07	99.42
19	9907	99.54	99.07	99.07	99.19
20	99.07	100.00	99.07	99.54	99.42
21	99.54	100.00	100.00	99.54	99.77
22	100.00	100.00	100.00	100.00	100.00
23	100.00	100.00	100.00	100.00	100.00
24	99.54	99.54	99.54	100.00	99.65
Totals	99.81	99.61	99.65	99.48	
Grand total over all times of day & over all participants					99.64

Mean Percentage Accuracy Rates at Each Time-of-Day for Each Participant and Over All Participants/Exp. 4

Participant Number	*Mean % Accuarcy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @ 1800*	*Total across all TOD for Each Participant*
1	93.52	97.22	97.22	99.07	96.76
2	91.20	96.30	98.15	99.07	96.18
3	97.22	87.96	92.59	96.76	93.63
4	99.54	98.61	99.54	99.54	99.31
5	99.54	99.07	99.07	99.54	99.31
6	98.15	98.61	93.06	97.22	96.76
7	99.54	98.61	99.07	96.76	98.50
8	99.54	98.15	99.54	99.54	99.19
9	99.54	99.54	97.69	97.22	98.50
10	100.00	98.61	98.61	100.00	99.31
11	97.22	97.22	97.22	98.15	97.45
12	99.07	94.44	96.76	99.07	97.34
13	99.54	99.07	99.54	97.22	98.84
14	100.00	100.00	100.00	100.00	100.00
15	97.22	96.76	98.15	91.67	95.95
16	97.69	97.69	95.37	98.15	97.22
17	93.98	96.30	96.76	96.76	95.95
18	99.07	100.00	100.00	98.61	99.42
19	97.69	90.28	93.98	94.44	94.10
20	96.30	94.44	96.76	99.07	96.64
21	95.37	94.91	88.89	91.20	92.59
22	99.07	99.54	100.00	99.54	99.54
23	99.07	98.15	98.15	93.98	97.34
24	95.83	97.22	97.22	98.15	97.11
Totals	97.70	97.03	97.22	97.53	
Grand total over all times of day & over all participants					97.37

Appendix 4: Experiment 5

(a): ANOVA Summary Tables for Experiment 5

Repeated Measures ANOVA Summary Table/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	331830.16	3	110610.05	69	5.17	0.003
Concreteness	6917.82	1	6917.82	23	0.97	0.336
TOD*Concreteness	12526.80	3	4175.60	69	0.88	0.456
Complexity	1336271.90	1	1336271.90	23	95.68	0.000
TOD*Complexity	19998.36	3	6666.12	69	1.50	0.223
Block	3182.74	2	1591.37	46	0.17	0.848
TOD*Block	61167.16	6	10194.53	138	1.67	0.133
Concreteness*Complexity	47111.71	1	47111.71	23	5.20	0.032
TOD*Concreteness* Complexity	24648.12	3	8216.04	69	1.61	0.194
Concreteness*Block	7076.32	2	3538.16	46	0.86	0.429
TOD*Concreteness*Block	15198.68	6	2533.11	138	0.60	0.735
Complexity*Block	15327.76	2	7663.88	46	2.52	0.092
TOD*Complexity*Block	43728.26	6	7288.04	138	1.39	0.224
Concreteness*Complexity *Block	16521.95	2	8260.98	46	1.85	0.169
TOD*Concreteness* Complexity*Block	11148.09	6	1858.01	138	0.36	0.903

Repeated Measures ANOVA Summary Table/Time-of-Day/Temperature

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	27.03	3	9.01	69	69.69	0.000
Before/After	7.20	1	7.20	23	76.32	0.000
TOD*Before/After	0.22	3	0.07	69	2.23	0.092

(b): Simple Interaction Effects/Simple Simple Main Effects/Simple Main Effects Summary Tables for Experiment 5

Simple Main Effect Summary Table (Complexity*Concreteness)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Complexity*Abstract Icons	942598.27	1	942598.27	23	51.98	0.000
Complexity*Concrete Icons	440785.34	1	440785.34	23	90.09	0.000

(c): Mean Percentage Accuracy Summary Tables for Experiment 5

Mean Percentage Accuracy Rates at Each Time-of-Day for Each Participant and Over All Participants

Participant Number	*Mean % Accuracy @0900*	*Mean % Accuracy @1200*	*Mean % Accuracy @1500*	*Mean %Accuracy @ 1800*	*Total Mean % Accuracy across all TOD for Each Participant*
1	99.54	100.00	99.07	99.54	99.54
2	97.22	98.15	96.76	99.54	97.92
3	100.00	100.00	100.00	99.54	99.88
4	99.54	99.07	99.07	100.00	99.42
5	100.00	99.54	99.54	99.54	99.65
6	99.07	100.00	99.07	99.54	99.42
7	100.00	100.00	100.00	99.07	99.77
8	100.00	100.00	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00
10	99.07	99.07	99.07	100.00	99.31
11	97.69	93.98	91.20	93.98	94.21
12	100.00	99.54	99.54	100.00	99.77
13	99.07	99.54	98.61	100.00	99.31
14	100.00	99.07	99.54	99.54	99.54
15	99.07	99.54	99.54	99.54	99.42
16	98.61	99.07	99.07	98.61	98.84
17	99.54	100.00	99.54	99.54	99.65
18	100.00	99.54	99.54	100.00	99.77
19	100.00	100.00	100.00	100.00	100.00
20	100.00	99.54	100.00	99.07	99.65
21	98.15	99.07	98.61	98.61	98.61
22	100.00	99.07	99.07	99.07	99.31
23	100.00	100.00	100.00	100.00	100.00
24	98.15	98.15	99.07	99.54	98.73
Totals	99.36	99.25	99.00	99.34	
Grand total over all times of day & over all participants					99.24

Appendix 5:Experiment 6

(a): ANOVA Summary Tables for Experiment 6

Repeated Measures ANOVA Summary Table/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	1877561.90	3	625853.97	69	2.99	0.03
Concreteness	22719.51	1	22719.51	23	0.20	0.66
TOD*Concreteness	118660.07	3	39553.36	69	0.88	0.45
Complexity	152344.77	1	152344.77	23	1.24	0.27
TOD*Complexity	46175.24	3	15391.75	69	0.26	0.85
Presence	210833843.80	1	210833843.80	23	128.33	0.00
TOD*Presence	7319.37	3	2439.79	69	0.04	0.99
Block	103680.93	3	34560.31	69	0.39	0.76
TOD*Block	610401.58	9	67822.40	207	0.97	0.46
Concreteness*Complexity	165193.31	1	165193.31	23	1.52	0.23
TOD*Concreteness*Complexity	104710.81	3	34903.60	69	0.66	0.57
Concreteness*Presence	235633.15	1	235633.15	23	6.41	0.01
TOD*Concreteness*Presence	48657.14	3	16219.05	69	0.47	0.70
Complexity*Presence	43918.31	1	43918.31	23	1.04	0.31
TOD*Complexity*Presence	192751.25	3	64250.42	69	2.37	0.07
Concreteness*Complexity*Presence	8021.84	1	8021.84	23	0.11	0.74
TOD*Concreteness*Complexity*Presence	200716.37	3	66905.46	69	1.64	0.18
Concreteness*Block	35741.43	3	11913.81	69	0.26	0.85
TOD*Concreteness*Block	671323.60	9	74591.51	207	1.83	0.06
Complexity*Block	315602.45	3	105200.82	69	2.03	0.11
TOD*Complexity*Block	423660.39	9	47073.38	207	0.85	0.56
Concreteness*Complexity*Block	382555.64	3	127518.55	69	3.02	0.03
TOD*Concreteness*Complexity*Block	419706.61	9	46634.07	207	1.06	0.39
Presence*Block	300023.70	3	100007.90	69	2.29	0.08
TOD*Presence*Block	354837.89	9	39426.43	207	0.92	0.51
Concreteness*Presence*Block	106974.48	3	35658.16	69	0.85	0.47
TOD*Concreteness*Presence*Block	290347.90	9	32260.88	207	0.94	0.49
Complexity*Presence*Block	73507.99	3	24502.67	69	0.68	0.56
TOD*Complexity*Presence*Block	354106.85	9	39345.21	207	1.03	0.41
Concreteness*Complexity*Presence*Block	279939.97	3	93313.32	69	2.31	0.08
TOD*Concreteness*Complexity*Presence*Block	326743.36	9	36304.82	207	0.92	0.50

Repeated Measures ANOVA Summary Table/Time-of-Day/Temperatures

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
TOD	15.13	3	5.04	69	115.63	0.000
Temperature Before/After	1.60	1	1.60	23	78.69	0.000
TOD*Temperature Before/After	0.03	3	0.09	69	1.24	0.301

(b): Simple Interaction Effects/Simple Simple Main Effects/Simple Main Effects Summary Tables for Experiment 6

Simple Main Effects Summary Table (Concreteness x Presence)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Abstract Icons*Presence	86647.08	1	86647.08	23	0.30	0.591
Concrete Icons*Presence	1151494.38	1	1151494.38	23	7.85	0.010

Simple Simple Main Effects Summary Table (Concreteness x Complexity x Block)/Time-of-Day/Response Times

	SS	*df*	*MS*	*Error*	*F*	*p*
W/S Factor(s)						
Concreteness*Simple Icons*Block 1	94749.53	1	94749.53	23	4.14	0.054
Concreteness*Simple Icons*Block 2	121787.99	1	121787.99	23	0.90	0.352
Concreteness*Simple Icons*Block 3	60596.98	1	60596.98	23	3.82	0.063
Concreteness*Simple Icons*Block 4	39649.81	1	39649.81	23	3.26	0.084
Concreteness*Complex Icons*Block 1	126252.61	1	126252.61	23	1.07	0.311
Concreteness*Complex Icons*Block 2	29521.22	1	29521.22	23	0.43	0.518
Concreteness*Complex Icons*Block 3	24652.06	1	24652.06	23	0.35	0.562
Concreteness*Complex Icons*Block 4	226087.68	1	226087.68	23	4.67	0.041

(c): Mean Percentage Accuracy Summary Tables For Experiment 6

Mean Percentage Accuracy Rates at Each Time-of-Day for all Participants

Participant No.	*Mean % Accuracy @ 0900*	*Mean % Accuracy @ 1200*	*Mean % Accuracy @ 1500*	*Mean % Accuracy @ 1800*	*Total Across all TOD for Each Participant*
1	97.92	96.99	96.99	96.30	97.05
2	96.06	97.69	96.53	97.22	96.88
3	94.44	92.59	96.30	94.68	94.50
4	97.69	95.14	95.37	92.36	95.14
5	96.76	98.15	98.15	98.15	97.80
6	97.69	97.92	96.99	97.22	97.46
7	94.91	93.52	93.29	95.14	94.22
8	97.69	98.15	97.69	96.06	97.40
9	95.83	96.30	96.99	96.53	96.41
10	92.82	90.28	92.13	96.06	92.82
11	94.68	93.06	88.66	93.06	92.37
12	91.67	90.97	87.73	88.66	89.76
13	93.98	94.91	94.44	91.90	93.81
14	95.37	91.90	94.68	96.53	94.62
15	94.21	96.53	94.44	92.82	94.50
16	95.83	96.99	99.07	97.45	97.34
17	95.37	96.30	98.15	95.83	96.41
18	96.99	95.60	97.45	95.14	96.30
19	95.60	95.60	96.06	95.60	95.72
20	94.68	95.14	90.28	90.97	92.77
21	94.91	94.91	94.68	95.60	95.03
22	99.07	98.15	98.84	98.15	98.55
23	95.83	95.14	96.53	95.83	95.83
24	98.38	97.92	97.69	98.38	98.09
Totals	95.77	95.41	95.38	95.24	
Grand total over all times of day & over all participants					95.45